

Causation & Explanation

Stathis Psillos

Central
Problems of
Philosophy
General Editor
John Shand

Causation and Explanation

Central Problems of Philosophy

Series Editor: John Shand

This series of books presents concise, clear, and rigorous analyses of the core problems that preoccupy philosophers across all approaches to the discipline. Each book encapsulates the essential arguments and debates, providing an authoritative guide to the subject while also introducing original perspectives. This series of books by an international team of authors aims to cover those fundamental topics that, taken together, constitute the full breadth of philosophy.

Published titles

Free Will

Graham McFee

Causation and Explanation

Stathis Psillos

Knowledge

Michael Welbourne

Ontology

Dale Jacquette

Relativism

Paul O'Grady

Scepticism

Neil Gascoigne

Truth

Pascal Engel

Universals

J. P. Moreland

Forthcoming titles

Action

Rowland Stout

Analysis

Michael Beaney

Artificial Intelligence

Matthew Elton & Michael Wheeler

Meaning

David Cooper

Mind and Body

Robert Kirk

Modality

Joseph Melia

Paradox

Doris Olin

Perception

Barry Maund

Rights

Jonathan Gorman

Self

Stephen Burwood

Value

Chris Cherry

Causation and Explanation

Stathis Psillos



© Stathis Psillos, 2002

This book is copyright under the Berne Convention.
No reproduction without permission.
All rights reserved.

First published in 2002 by Acumen
Reprinted 2009

Acumen Publishing Limited
Stocksfield Hall
Stocksfield
NE43 7TN
www.acumenpublishing.co.uk

ISBN: 978-1-902683-41-6 (hardcover)
ISBN: 978-1-902683-42-3 (paperback)

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from
the British Library.

Printed in the UK by the MPG Books Group.

For my daughter *Demetra*

Contents

<i>Preface</i>	ix
<i>Abbreviations</i>	xiii
Introduction	1
I Causation	
1 Hume on causation	19
2 Regularities and singular causation	57
3 Causation and counterfactuals	81
4 Causation and mechanism	107
II Laws of nature	
5 The regularity view of laws	137
6 Laws as relations among universals	159
7 Alternative approaches to laws	179
III Explanation	
8 Deductive–nomological explanation	215
9 Statistical explanation	241
10 Explanation of laws	263
11 The metaphysics of explanation	281
<i>Notes</i>	295
<i>References</i>	309
<i>Index</i>	317



Preface

When I first started thinking about this book, I intended to write a short introduction to the philosophical debates surrounding the concepts of causation and explanation. Part of my motivation was the thought that, although there are quite a few splendid books on either causation or explanation, there was no book treating these two subjects together and aiming to cast light on their connections. As time passed, the short introduction grew bigger and bigger. In fact, it dawned on me that I couldn't adequately address the *link* between causation and explanation, unless I also wrote extensively about the laws of nature. And that's what I did. As a result, the title of the book should have been *Causation, Laws and Explanation*. In the end, the word *laws* was left out, yet the middle part of the book (Laws of Nature) forms its backbone.

The book is broad in scope, but by no means comprehensive. It aims to introduce students to the *main* theories of causation, laws and explanation. But it also ventures into more recent developments and approaches in these areas, aiming to show that, far from being philosophically sterile, these topics are very much alive and exciting. The book presupposes little knowledge of either metaphysics or philosophy of science and can be used in intermediate and advanced undergraduate courses. Yet I very much hope that professional philosophers, even specialists in these areas, will find it useful. The book presents no new theory of either causation or laws or explanation, but it does try to evaluate, critically discuss and draw connections among existing approaches, including some very recent ones. Having the firm belief that a book aimed to be a textbook should *not*

be partisan, I have tried to be fair in my assessment of the views I discuss. Yet I make no secret of my own view that, suitably understood, the regularity view of causation, the regularity view of laws and a nomological approach to explanation are still serious philosophical contenders.

During the early stages of my work for this book, I had the privilege of extended discussions with Wes Salmon. Wes had promised me to make detailed comments on the final draft, but his untimely death deprived me of this great honour (and the philosophical community of a first-rate philosopher and a real gentleman). I am sure the book would have been much better had Wes looked at it. Another great source of inspiration (as for very many others) has been David Lewis's work. All those who work on causation, laws and explanation will feel his untimely loss. I owe a great intellectual debt to all philosophers whose views I discuss in the book, but I feel that I have to make a special reference to the inspiration I got from the work of David Armstrong, Nancy Cartwright, John Earman, Carl Hempel, Philip Kitcher, J. L. Mackie, D. H. Mellor, Peter Railton, F. P. Ramsey and Barry Stroud. Once more, David Papineau has been a great teacher for me. His encouragement has been instrumental for the completion of the book. Two anonymous readers for *Acumen* have made important suggestions and comments, which, I hope, have led to a substantial improvement of the final product. Their hearty endorsement of the draft they read gave me the strength I needed to carry on. Some other colleagues and friends gave me thorough comments on several portions of the book. I should particularly like to thank Helen Beebe, Alexander Bird, Mark Lange and Rebecca Schweder. The graduate students who attended a course based on this book last year were wonderful critics of both its content and its style. And my colleagues in the Department of Philosophy and History of Science at the University of Athens created a very friendly environment for me to work in. The errors that, I am sure, still remain would have been many more without the generous help I received from all those people. Steven Gerrard at *Acumen* deserves special mention for his support throughout the completion of this book. Many thanks should also go to Jenny Roberts, who copy-edited this book, for the care with which she read the typescript.

A good part of this book was written during the summer of 2001 in Platania, a beautiful village at the foot of Mount Idi in Crete. I am

indebted to my family and other local people for making my stay there so comfortable and for giving me the space I needed to work. My wife Athena has been a source of endless care and love (as well as a great teacher of how psychologists think about causation). Finally, a couple of months after the book had been sent to the readers, our daughter, Demetra, was born. I couldn't wait to dedicate it to her.

S. P.

Athens, May 2002

Abbreviations

ADT	Armstrong–Dretske–Tooley
BT	the basic thesis: all causal explanations of singular events can be captured by the deductive–nomological model
CE	the thesis that all explanation of singular events is causal explanation
CI	causal interaction
CQ	conserved quantity
DN	deductive–nomological
DNP	deductive–nomological–probabilistic
DS	deductive–statistical
HNS	Humean nomic supervenience
HS	Humean supervenience
IP	inductive probabilism
IS	inductive–statistical
IT	insufficiency thesis
MRL	Mill–Ramsey–Lewis
MT	mark-transmission
NT	the set of all laws
PLV	the Principle of Limited Variety
PUN	the Principle of Uniformity of Nature
RMS	the Requirement of Maximal Specificity
RVC	the Regularity View of Causation
RVL	the Regularity View of Laws
SR	statistical-relevance

Introduction

The birth of our daughter was the *cause* of great happiness to my wife and me. This *explains* why I decided to dedicate this book to her. It also *caused* certain changes in our life (for instance, that our study at home had to be converted to a nursery). It *brought about* a delay in the completion of the current book, which (hopefully) *explains* why this book might well be a bit better than it would have been had I rushed to finish it. It is *because of* her birth that I have come to realize how challenging and exciting parenthood is. And this *explains* my recent interest in books about babies. And so on and so forth.

Causal and explanatory talk is so pervasive in our everyday life, as well as in the sciences, that its importance can hardly be exaggerated. We search for causes and look for explanations in order to *understand* how and why things around us are the way they are, or behave and change in certain ways. But we also search for causes and look for explanations in order to *intervene* in the course of nature (or in the course of events, in general) and bring certain effects about or prevent others from occurring. We are interested in causation and explanation because we are *thinkers* and *agents*, because we are both theoretical and practical beings. We are worried, for instance, about the future because of certain recent developments (e.g., the destruction of the environment, or the revival of terrorism and of war, or the depletion of natural resources, or the resurgence of racism and xenophobia). We think (rightly) that we have identified at least some of the causes of this worry and we want to act to eliminate them, as well as their own causes. Theory

and practice are cemented together by the chains of causation. We offer *reasons* for our actions. But reasons, as Donald Davidson has famously stressed, can be causes of action. Besides, learning the causes of an event (be it the breaking out of the Second World War, or the collapse of a bridge, or the famine in Ethiopia) enhances our knowledge of why it happened as well as our ability to prevent similar events from happening. It also gives us *reasons* to form well-supported beliefs.

It is, of course, one thing to acknowledge, or stress, the centrality of the concepts of causation and explanation in our intellectual and practical life, and quite another thing to say what these concepts are concepts *of*. What is causation and what explanation? And how exactly are the two related? Answering these questions is the job of philosophers. Intuitively, explanation and causation go hand-in-hand. Isn't it a platitude, after all, that in order to explain something, you need to cite its causes? This platitude might not be fully general, since there are non-causal explanations (most typically, mathematical explanations), but it seems to go a long way in highlighting the *link* between causation and explanation: causes do explain and explanation does proceed via stating causes. But can we go beyond this platitude? In particular, can we offer adequate theories of causation and explanation? Can we specify the semantics of causal and of explanatory talk?

Such questions become even more urgent if we take into account that, as of late, philosophers seem to make free use of the concepts of causation and explanation in their attempts to analyse and solve traditional philosophical problems. We now have causal theories of *knowledge*, causal theories of *perception*, causal theories of *reference*, causal theories of *identity through time*, causal-role theories of *mental states* and so on. All these are piled up on top of more traditional problems such as the problem of *mental causation* (how, that is, the mental can act causally on the physical), or the problem of what exactly are the *relata* of (i.e. the things that are related by) causal relations (events, facts, states of affairs, etc.), or the problem of the *explanatory autonomy* of the special sciences, or the nature of *inference to the best explanation*. It seems either that we have to appeal to some shaky prephilosophical intuitions about causation and explanation when we investigate all of the above, or else that we need to do some serious groundwork to clarify what exactly we

refer to when we speak of causation and explanation. A central aim of this book is to present and critically discuss some of this ground-work in an attempt to clarify some of the basic conceptual issues that are involved in the philosophical debates about causation, laws of nature and explanation.

Philosophers have long disagreed about the nature of causation and explanation. They have offered different theories, either within the same or within rival metaphysical agendas. In Michael Scriven's (1975: 3) apt words, the concepts of causation and explanation "enjoy a curious love-hate relationship" with philosophers. Most think that these concepts are central to all our thinking about science (as well as about our everyday affairs) and try hard to analyse them, but there are some who *deny* their importance and suggest that the sooner they are fully dispensed with the better.

Causation

Let's start with causation. Perhaps the most famous denier of causation was Bertrand Russell (1918), who actually thought that the concept of causation was incoherent. But this was just as well for him, since, as he alleged, physics has stopped looking for causes: for "there are no such things". Here is his famous dictum: "The law of causality, I believe, like much that passes muster among philosophers, is a relic of a bygone age, surviving, like the monarchy, only because it is erroneously supposed to do no harm."¹ Now, even if Russell were right about physics – although what he asserts with confidence is, to say the least, debatable – he is definitely wrong about the other sciences. Even a cursory look at subjects such as economics, psychology and biology will persuade the non-believer that scientists do hunt for causes and assume that causes are there to be hunted. Quite powerful methods (such as randomized controlled experiments – *aka* clinical trials – and causal modelling) have been developed to aid the discovery of causal dependencies between factors and magnitudes. Far from having survived because of the erroneous supposition that it does no harm, the search for causes has been both successful and beneficial.

In a fashion similar to Russell's, Rudolf Carnap also noted that, strictly speaking, the concepts of "cause" and "effect" are meaningful only within the "perceptual world", and that, having as their

domain of application that world, “they are infected with the imprecision which attaches to concept formations within this world” (1928: 264). This may well be so, but all that it implies is that the concept of causation needs analysis and regimentation. As a matter of fact, the project initiated by Carnap and his fellow logical empiricists (but also followed by other eminent philosophers) was an attempt to *characterize* and *rescue* a legitimate core from the concept of causation, by equating causation with *de facto* invariable succession or actual regularity. This is what has come to be known as the Regularity View of Causation (RVC). It is typically seen as offering a *reductive* account of causation. As with all reductive accounts, causal talk becomes legitimate, but it does not imply the existence of a special realm of causal facts that make causal talk true, since its truth conditions are specified in non-causal terms, that is, in terms of spatiotemporal relations and actual regularities.

Most of the empiricists’ ammunition has come from Hume’s critique of causation. Ever since David Hume’s work, philosophers of an empiricist persuasion have thought that the concept of causation is too mysterious or metaphysical to be taken seriously without any further analysis. Hence, they engaged in an attempt to *demystify* causation. They thought that the main culprit was the idea that causation implies the existence of *necessary connections* in nature, that is, connections between the causally related events that make it, somehow, necessary (or inescapable) that the effect follows from the cause. Hume was taken as the great denier of such necessary connections and as the one who conclusively showed that there were no such things to be found in nature. This denial of necessary connections in nature may be seen as the hallmark of modern Humeanism.

Some Humeans (most notably John Stuart Mill and John L. Mackie) advanced more sophisticated versions of RVC. A prominent thought has been that causation should be analysed in terms of *sufficient and necessary conditions* (roughly, an event *c* causes an event *e* if and only if (iff) there are event-types *C* and *E* such that *C* is necessary and sufficient for *E*). Another one has been that to call an event *c* the cause of an event *e* is to say that there are event-types *C* and *E* such that *C* is an *insufficient* but *necessary* part of an *unnecessary* but *sufficient* condition for *E* – *aka inus* condition). A rather important objection to Humeanism has been that regularity is not

sufficient for causation. There are too many regularities in nature and not all of them are, intuitively, causal. So Humeans have been inequalitarians towards regularities. They have tried to characterize the kind of regularity that can underpin causal relations by tying causation to laws of nature. However, other philosophers who advocate Humeanism downplay the role of regularities (or laws) in causation. A rather prominent approach has been Lewis's account of causation in terms of *counterfactual conditionals* (roughly, an event *c* causes an event *e* iff if *c* hadn't happened then *e* wouldn't have happened either). To be sure, regularities do enter the counterfactual approach to causation but in a roundabout way: as means to capture the conditions under which counterfactual assertions are true.

Many non-Humean theories deny forcefully that the analysis of causation need involve regularities, either directly or indirectly. John Curt Ducasse's *single-difference* account (roughly that an event *c* causes an event *e* iff *c* was the last – or the only – difference in *e*'s environment before *e* occurred) takes causation to link individual events independently of any regular association that there may or may not be between events like the cause and events like the effect. Salmon's *mechanistic* approach (roughly that an event *c* causes an event *e* iff there is a causal process that connects *c* and *e*) stresses that there is a local tie between a particular cause and a particular effect. Causation, non-Humeans argue, is essentially singular: a matter of *this* causing *that*.

Some philosophers think, contra Hume, that causation is directly *observable*. Others take it to be a *theoretical* relation, posited to explain a particularly robust connection between some events. Many philosophers think that if we are to avoid methodological obscurity and conceptual circularity, we have to cast the analysis of causation in non-causal terms. However, others argue that such an analysis is impossible. They dispel the charge of circularity by arguing that the concept of causation is so *basic* that it cannot be really analysed in non-causal terms. All that can be done, they claim, is to offer an enlightening account of how several causal concepts are interlinked and mutually understood.

I shall now try to offer a brief diagnosis as to why there is no general agreement among philosophers on what causation is. In a nutshell, the diagnosis is that the concept of causation seems to be

characterized by conflicting intuitions, which, although almost equally central, cannot be all accommodated in a single theory. A number of philosophers – most notably D. H. Mellor (1995), Peter Menzies (1996) and David Armstrong (1999) – have recently tried to approach causation via what they have called “platitudes”.² Perhaps, it’s best to make a distinction between *platitudes* (assuming that there are some platitudinous features of causation that any theory should accommodate) and *intuitions* (assuming that there are some firm prephilosophical views about what causation is). Some of the platitudes of causation are these:

- The *difference* platitude: causes make a difference – that is, things would be different if the causes of some effects were absent. This platitude is normally cast in two ways: the *counterfactual* way – if the cause hadn’t been, the effect wouldn’t have been either; and the *probabilistic* way – causes raise the *chances* of their effects – that is, the probability that a certain event happens is higher if we take into account its cause than if we don’t.
- The *recipe* platitude: causes are recipes for producing or preventing their effects – that is, causes are the means to produce (or prevent) certain ends (effects).³ This platitude is normally cast in terms of *manipulability*: causes can be manipulated to bring about certain effects.
- The *explanation* platitude: causes explain their effects, but not vice versa.
- The *evidence* platitude: causes are evidence for their effects – that is, knowing that *c* causes *e*, and knowing that *c* occurred, gives us (some) reason to expect that *e* will occur.

It’s not hard to agree that each and every theory of causation should accommodate these platitudes, that is, show how each of them is brought out by whatever, according to the theory, constitutes the relation of cause and effect. But there are two central intuitions about causation that also need to be taken into account.

- The *intrinsic-relation* intuition: whether or not a sequence of two distinct events *c* and *e* is causal depends wholly on the events *c* and *e* and their own properties and relations, that is, it

depends wholly on the intrinsic and local features of the actual sequence of events. For instance, according to this intuition, when we say that *the hitting with the hammer caused the smashing of the porcelain vase* what makes our assertion true has only to do with the properties of the particular hammer, the particular vase and the particular hitting.

- The *regularity* intuition: whether or not a sequence of two distinct events *c* and *e* is causal depends on whether or not events like *c* are regularly followed by events like *e*. This intuition is captured by the dictum “same cause, same effect” and is underpinned by an epistemic consideration; namely, that we are unwilling to pronounce a sequence of events *c* and *e* causal unless there has been a regular association between events like *c* and events like *e*. For instance, according to this intuition, when we say that *the hitting with the hammer caused the smashing of the porcelain vase* what makes our assertion true has to do with the fact that the hitting of porcelain vases with hammers is regularly followed by the smashing of the vases.⁴

Now, these two intuitions pull in contrary directions. The regularity intuition implies that a sequence of events is causal if and only if it instantiates a regularity. Hence, it implies that the relation of cause and effect is *extrinsic* to its relata. It makes causation dependent on *general* facts: on what happens at other places and at other times. The intrinsic-relation intuition opposes all this. It takes causation to be wholly dependent on *singular* facts: on what happens there and then, in the actual sequence of events, independently of any regularities. It would be a daunting (not to say just outright impossible) task to advance a theory that respects both of these intuitions. Most typically, Humeans base their theories on the regularity intuition, while non-Humeans base theirs on the intrinsic-relation one. A somewhat detailed investigation of the distinction between Humean and non-Humean approaches has to wait until the end of Chapter 4 (section 4.5), where, after we have examined several accounts of causation, we shall offer a map of the terrain.

It would do no harm, however, to highlight three dimensions along which the discussion about causation can be based. We have already seen the first two. The first concerns the distinction

between generalist and singularist theories. The second dimension concerns the distinction between theories that aim at an extrinsic characterization of causal relations and theories that go for an intrinsic one. The third dimension concerns the distinction between reductive approaches and non-reductive ones. Reductive approaches argue that causation is dependent on (some say it *supervenes* on) non-causal features of the world (e.g. regularities), while non-reductive accounts take causation to be ontically autonomous: an irreducible relation among events. On a first approximation, then, one could say that Humean accounts of causation take the first sides of the three distinctions: they are generalist, extrinsic and reductive. And non-Humean accounts take at least one of the remaining sides of the three distinctions: they are singularist or intrinsic or non-reductive. As further investigation will show, however, things are more complicated. Perhaps it's not very profitable to try to divide theories of causation sharply into Humean and non-Humeans, although, as we shall see, we can go some way towards achieving this task.

Laws of nature

Most Humeans have come to adopt what may be called the Regularity View of Laws (RVL): laws of nature are regularities. However, they have a hurdle to jump. For not all regularities are causal. Nor can all of them be deemed laws of nature. The night always follows the day, but it is *not* caused by the day. Nor is it a law of nature that all coins in my pocket are Euros, although it is a regularity. So the Humeans have to draw a distinction between the good regularities (those that constitute the laws of nature) and the bad ones (those that are merely accidental). Only the former, it was thought, can underpin causation and play a role in explanation. We shall see in some detail the various empiricist attempts to draw this distinction, and in particular, what I shall call the *web-of-laws* view.

According to this view, the regularities that constitute the laws of nature are those that are expressed by the axioms and theorems of an ideal deductive system of our knowledge of the world, and, in particular, of a deductive system that strikes the *best* balance between simplicity and strength. Simplicity is required because it disallows extraneous elements from the system of laws. Strength is

required because the deductive system should be as informative as possible about the laws that hold in the world. Whatever regularity is not part of this *best system* is merely accidental: it fails to be a genuine law of nature. The gist of this approach, which has been advocated by Mill, Ramsey and Lewis, is that no regularity, taken in isolation, can be deemed a law of nature. The regularities that constitute laws of nature are determined in a kind of holistic fashion by being parts of a structure. As we shall see, the *web-of-laws* view does succeed, to a large extent, in answering the question “What is a law of nature?”. Yet its critics argue that it compromises the fully objective status that the laws of nature are, typically, taken to have. Why, they tend to ask, should lawhood have anything to do with how our *knowledge* of the world is organized in a deductive system? There is no doubt that the Humeans should try to dispel this charge of subjectivity. The good news, however, is that they can, to some extent at least, secure the objectivity of laws. But, as I shall argue, in order to do so they have to adopt a certain metaphysical picture; namely, that the world has an *objective* nomological structure. This structure, to be sure, will be a structure of *regularities*. Yet it may well be the case that it is objective relations among the elements of this structure, and *not* our beliefs about them, that determine what regularities are parts of this structure, and hence what regularities constitute laws of nature.

It should be noted, however, that even if the *web-of-laws* view can be deemed successful, there is a price to pay. By denying that there is any necessity in causation, the Humeans have to deny that there is any necessity in the laws of nature. Their non-Humean opponents then are quick to point out that without some appeal to a sufficiently strong concept of necessity, the distinction between laws of nature and accidental regularities will not be robust enough to support either causation or explanation. What, in short, emerges from their arguments is the view that lawhood cannot be reduced to regularity (not even to regularity-plus-something-that-distinguishes-between-laws-and-accidents). Lawhood, we are told, is a certain necessitating relation among properties (*universals*). It is noteworthy that *both* the Humeans and the advocates of the view that laws embody necessitating relations among properties agree that laws of nature are *contingent*. They do not hold in all possible worlds: they could be different, or there could be no laws at all. Yet, there has been a

growing tendency among non-Humeans to take laws of nature to be metaphysically necessary. A standard Humean retort to all these views is that, far from being enlightening, the notion of necessitation that is supposed to characterize the laws of nature (either as a contingent relation among properties or as embodying a stronger metaphysical sense of necessity) is wrong-headed and obscure.

Explanation

When it comes to the concept of explanation things may seem more promising. Here, the Logical Empiricist project of demystifying causation culminated in the attempts made by Hempel and his followers to analyse the concept of causation in terms of the concept of *explanation*. They thought that the latter could be made to be scientifically respectable by being itself analysed in terms of the concept of laws of nature and the concept of a deductive argument. The famous (to some notorious) deductive–nomological (DN) model of explanation has been a systematic attempt to subsume causation under causal explanation and to show that the latter can be fully understood and legitimized as a species of deductive argument, with one of its premises stating a universal law. In fact, the empiricist project was advanced further by enlarging the kind of arguments that can be explanations so as to include *inductive* arguments (and statistical, as opposed to universal, laws).

This reliance on laws makes it very pressing for the advocates of the DN model to have a neat way to distinguish between genuine laws of nature and accidentally true generalizations, for it is only the former that can be mentioned in legitimate explanations. The presence of an accidental generalization as a premise in a DN argument would amount to a cancellation of the nomological side of the argument. One can certainly deduce (a description of) the fact that this apple is ripe from the general statement “All apples in the fruit bowl are ripe” and the premise “this apple is in the fruit bowl”. Yet this deduction hardly explains *why* this apple is ripe. Compare, however, the above with the following case. Intuitively, at least, the following argument is a perfectly legitimate explanation of the fact that Pluto describes an ellipse: Pluto is a planet and all planets move in ellipses. The difference, the advocate of the DN model would argue, is that *All apples are in the fruit bowl* is an accident, whereas

All planets move in ellipses is a genuine law. As a result of all this, the project of developing an adequate DN model of explanation can proceed only hand-in-hand with an attempt to characterize the genuine laws of nature.

The irony of the empiricist project is that what came out of the front door seemed to be re-entering from the window. For it seems that we cannot distinguish between good and bad explanations of some phenomena, *unless* we first distinguish between causal and non-causal explanations, or better between those explanations that reflect the causal connections between what-is-doing-the-explaining (the *explanans*) and what-is-explained (the *explanandum*) and those that do not. So it seems that we first need to sort out the concept of causation and then talk about causal explanation. If this is right, then the empiricist project outlined above gets things the wrong way around.

Yet there are plausible ways for modern empiricists to argue that, suitably understood in terms of the concept of *unification*, explanatory relations can still subsume causal relations under them. Put in a nutshell, the idea is that explanation proceeds via unification into a deductive system: a certain fact, or a regularity, is explained when a description of it is deduced within a unified deductive system. Causal relations, then, are said to mirror explanatory relations within an ideal unified deductive system. What is really interesting here is that the concept of unification can be connected with the *web-of-laws* view. Unification proceeds by minimizing the number of regularities that have to be accepted as *brute* (or as unexplained explainers). These regularities might well be accepted as the fundamental laws of nature and be captured by the axioms of an ideal deductive system that strikes the best balance between simplicity and strength. Such an ideal deductive system is none other than a unified deductive system. In line with the *web-of-laws* view, the fundamental laws of nature are the best unifiers. Yet those philosophers who resist the attempt to subsume causation under explanation point out that the foregoing view of explanation as unification will not deliver the goods. Not only is it possible that the world be disunified, but, more importantly, it seems that the foregoing view is unable to specify the conditions under which an explanation is correct. It seems, we are told, that we need to rely on the *causal structure of the world*; it is because the world has a certain causal

structure that some explanations are correct (those that capture this causal structure), while others are not. If these philosophers, notably Salmon, are right, then causal relations simply *cannot* mirror explanatory relations, even within an ideal unified system of the world. Rather, the opposite should be the case: explanatory relations, even within an ideal unified system of the world, should reflect (and capture) ontically prior causal relations.

In any case, not all philosophers agree that causal explanation should be tied to laws and have the form of an argument. Opponents of the DN model argue that explanation should rely on finding out the causes of the *explanandum*, but it need not cite laws: presenting information about the causal history of an event, or citing factors that raise the probability of an event to happen, or even stating some invariant relations among the *explanandum* and the *explanans* is taken to be enough for a good causal explanation.

The fact of the matter is that the concepts of causation, laws of nature and explanation form a quite tight web. Hardly any progress can be made in the elucidation of any of those without engaging in the elucidation of at least some of the others. All we may then hope for is not strict analysis, but some enlightening account of their interconnections.

The menu

Although I have already hinted at the contents of this book, a more orderly presentation of the chapters to follow may help the readers orientate themselves better. Chapter 1 is about Hume and the setting up of RVC. It unravels the two projects that Hume was engaged in; namely, the analysis of causation as it is in the world and the analysis of the nature of causal inference. It culminates with a discussion of Hume's two definitions of causation. It ends with a short discussion of recent re-interpretations of Hume's views, which distance Hume from RVC, and which have given rise to the *new Hume debate*. Chapter 2 discusses Mill's elaboration of RVC, with special reference to his *methods of agreement and difference* for discovering causal laws. It also examines Ducasse's attempt to mount a major challenge to RVC and to motivate a singularist approach to causation based on Mill's method of difference. It criticizes a popular argument to the effect that causation is an observable relation and ends with

some discussion of Davidson's attempt to reconcile the Humean and the singularist approach. Chapter 3 discusses two major attempts to analyse causation in terms of *counterfactual conditionals*; namely, Mackie's and Peter Lewis's. It also analyses Mackie's own formulation of RVC (which takes causes to be *inus* conditions). Finally, it ventures into a discussion of recent counterfactual approaches, such as Huw Price's and Menzies' human agency view and Daniel Hausman's and James Woodward's interventionist view. Chapter 4 (which concludes the part on causation) investigates theories that characterize the link between cause and effect in terms of some mechanism that *connects* them. After discussing views that argue that in the transition from the cause to the effect something *persists* or something gets *transferred*, it focuses on Salmon's early and later attempts to analyse physical causation in terms of causal processes and the transference of conserved quantities. It moves on to analyse Phil Dowe's attempt to analyse causation without an appeal to counterfactuals. It concludes with offering a rough *conceptual map* of the terrain of causal theories.

Part II of the book (on laws of nature) starts with Chapter 5, whose main aim is to critically discuss RVL. It starts with naïve versions of RVL, which simply equate laws with regularities, raises the issue of how RVL needs to be supplemented to account for the distinction between laws and accidents, and examines two major attempts towards such a supplementation: the view that the difference between laws and accidents is merely a difference in our epistemic attitudes towards them, and the much-promising Mill–Ramsey–Lewis (MRL) view, which takes laws to differ from accidents in that the regularities that are laws form a tight web. Chapter 6 focuses on non-Humean theories of lawhood. It analyses the view of Armstrong, Fred Dretske and Michael Tooley that laws are relations of contingent necessitations among properties. It strongly questions the notion of necessitation that they appeal to. It then moves on to discuss even stronger theories of lawhood, which take laws to be metaphysically necessary. It ends with a critical examination of recent arguments against the Humean view that laws supervene on non-nomic facts. Chapter 7 presents recent attempts to supersede the traditional framework of the debate on laws, by focusing more on methodological aspects of the role of laws. Among the issues that are being examined are Woodward's characterization

of laws as relations that are invariant-under-interventions, Cartwright's appeal to capacities, Lange's view on the collective stability of laws and Mellor's focus on the link between laws and natural properties. This chapter will end with a cost-benefit analysis of the major views of laws and will suggest that, on balance, RVL is still the best characterization of what a law of nature is.

The final part of the book (Part III) is on *explanation*. It starts, in Chapter 8, with the DN model of explanation. It highlights the Humean-empiricist project to deal with causation via the concept of explanation and critically discusses the counter-examples that were supposed to have uprooted the DN account. It tries to show with some precision what these counter-examples have and what they have *not* shown. It ends with an investigation of Lewis's theory of causal explanation. Chapter 9 extends the empiricist project to statistical explanation and discusses Hempel's inductive-statistical (IS) model and its problems, Salmon's statistical-relevance (SR) model and Railton's deductive-nomological-probabilistic (DNP) account. Chapter 10 extends further the empiricist project to the explanation of laws. It analyses Michael Friedman's and Philip Kitcher's models of explanatory unification. Finally, Chapter 11 engages with the issue of the connection between causation and explanation. Here, the challenge is whether the Humean-empiricist project can be completed – whether, that is, it can be shown that the explanatory relations are primary and that, somehow, the causal relations follow from them. It will be argued that Humeans can go a long way towards meeting this challenge, but that, in doing so, they have to adopt the realist view that the world has an objective structure, in which mind-independent regularities form a unified system. This insight, I will suggest, can be found in the work of Ramsey.

Absences

Three absences from this book require brief apology. I do not discuss (apart from a few brief mentions) issues related to *probabilistic causation*. We do rightly claim that, for instance, smoking causes lung cancer or that aspirin relieves headaches, even though there is no regular association (or deterministic connection) between smoking and lung cancer or taking aspirin and relief from headaches. Some philosophers think that this is already a good argument against the

view that causation is connected with invariable sequences or regularities. They then try to analyse causal claims in terms of probabilistic relations among variables, capitalizing on the intuition that causes (mostly, but not invariably) raise the probabilities of their effects. Some think that there are good empirical reasons to jettison determinism (roughly, the view that each and every event has a fully sufficient set of causes) in favour of indeterminism (roughly, the view that there are genuinely chancy events). They then try to show that indeterminism and causation mix well, given the thought that a certain event can be *caused* to happen even though its cause made only a difference to its chance of happening. Interestingly, these ideas are extended to deterministic causation as well, with the prime thought being that an effect is deterministically caused to happen if its probability, given its cause, is unity. It is also noteworthy that probabilistic theories of causation are advanced by both Humeans (who think that causal connections are reducible to relations of probabilistic dependence) and non-Humeans (who think that causal relations are *not* reducible to probabilistic relations but, nonetheless, take the latter to illuminate causation.) Discussing these intricate matters would have made this book unmanageably long. So the reader is advised to look at Patrick Suppes (1984), David Papineau (1985) and Ellery Eells (1991) for excellent accounts of probabilistic causation. For what it is worth, my own view is close to Hausman's (1998: 186). I too think that acceptance of indeterminism implies the acceptance of uncaused things, but that there can be fully deterministic causation of probabilistic states.

Another issue I do not discuss (apart from a few scattered observations and remarks) concerns the *direction of causation*. Why is it the case that causes precede their effects in time? Some philosophers (including Hume) thought that this feature is conceptually constitutive of causation, while others think that it is an empirical feature of the actual world, which needn't obtain in other possible worlds. Other philosophers try to define the order of causation independently of the concept of time, so that they can then explain the direction of time in terms of the direction of causation. All philosophers who have thought hard about causation have dealt with this issue of *causal priority*. But, here again, I would advise the interested reader to look at Paul Horwich (1987) and Hausman (1998) for excellent guidance into all this.

Finally, a third issue I do not touch at all relates to the so-called *pragmatics of explanation*. Some philosophers focus on the *act* or the *process* of explaining, instead of the *product* of explanation. They argue that an explanation should be seen as an answer to a why-question and note that the *relevant* answers will depend on the presuppositions or the interests of the questioner, on the space of alternatives, and, in general, on the context of the why-question. Here is Alan Garfinkel's (1981: 21) famous example. A priest asked Willie Sutton, when he was in prison, "Why did you rob banks?", to which Sutton replied, "Well, that's where the money is". Garfinkel's thought is that this is a perfectly legitimate answer for Sutton, because for him the space of relative alternatives (the contrast class) concerns robbing groceries or diners or petrol stations, and so on. But the space of relevant alternatives for the priest is quite different: not robbing anything, being honest, and so on. The difference of perspective can be brought out by placing the emphasis on different parts of the question: "Why did you rob *banks*?" as opposed to "why did you *rob* banks?" Pragmatic theories of explanation, very different in their details but quite similar in their overall focus on the act of explaining and the contrast classes, have also been offered by Bas van Fraassen (1980) and Peter Achinstein (1983).

With all this in mind, it's now time to leave the starter and move on to the main three-course meal. I hope you enjoy it.

I Causation

1 Hume on causation

1.1 The regularity view of causation

A good starting point for our philosophical endeavours is David Hume's account of causation. His work on this subject has been, by far, the most important and influential ever. Hume's account has been taken to be a *reductive* one. It's been typical to call this account the Regularity View of Causation (RVC).

RVC

c causes e iff

- (a) c is spatiotemporally contiguous to e ;
- (b) e succeeds c in time; and
- (c) all events of type C (i.e., events that are like c) are regularly followed by (or are constantly conjoined with) events of type E (i.e. events like e).

So, on RVC, causation reduces to spatiotemporal contiguity, succession and constant conjunction (regularity). It reduces, that is, to non-causal facts. A corollary of RVC is that there is no necessity in causation: there is no necessary connection between the cause c and the effect e that goes beyond – or underpins – their regular association. RVC has been espoused by many eminent philosophers and has been taken to be the official Humean view. Here are a few representative statements of it.

The Law of Causation . . . is but the familiar truth that invariability of succession is found by observation to obtain between

every fact in nature and some other fact which has preceded it . . . (Mill 1911: 213)

We must ask ourselves: when we assume causation, do we assume a specific relation, cause-and-effect, or do we merely assume invariable sequence? That is to say, when I assert “every event of class A causes an event of class B”, do I mean merely “every event of class A is followed by an event of class B”, or do I mean something more? Before Hume the latter view was always taken; since Hume, most empiricists have taken the former. (Russell 1948: 472)

In nature one thing just happens after another. Cause and effect have their place only in our imaginative arrangements and extensions of these primary facts. (Ayer 1963: 183)

The trouble with causation is, as Hume pointed out, that there is no evident way of distinguishing it from mere invariable succession. (Quine 1974: 5)

[a] statement about a causal relation . . . describes an observed regularity of nature, nothing more. (Carnap 1974: 201)

[According to Hume] to say of a particular event *a* that it caused another event *b* is to place these two events under two types, *A* and *B*, which we expect to be constantly conjoined in the future as they were in the past. (Kripke 1982: 67)

RVC has been traced to what Hume thought and said. Take, for instance, a famous passage from his *Abstract to A Treatise of Human Nature*, in which Hume discusses one of his favourite examples of causation, the collision of two billiard balls:

Here is a billiard-ball lying on the table, and another ball moving towards it with rapidity. They strike; and the ball, which was formerly at rest, now acquires a motion. This is as perfect an instance of the relation of cause and effect as any which we know, either by sensation or by reflection. Let us therefore examine it. 'Tis evident, that the two balls touched one another

before the motion was communicated, and that there was no interval betwixt the shock and the motion. *Contiguity* in time and place is therefore a requisite circumstance to the operation of all causes. 'Tis evident likewise, that the motion, which was the cause, is prior to the motion, which was the effect. *Priority* in time, is therefore another requisite circumstance in every cause. But this is not all. Let us try any other balls of the same kind in a like situation, and we shall always find, that the impulse of the one produces motion in the other. Here therefore is a *third* circumstance, *viz.*, that is a *constant conjunction* betwixt the cause and effect. Every object like the cause, produces always some object like the effect. Beyond these three circumstances of contiguity, priority, and constant conjunction, I can discover nothing in this cause. The first ball is in motion; touches the second; immediately the second is in motion: and when I try the experiment with the same or like balls, in the same or like circumstances, I find that upon the motion and touch of the one ball, motion always follows in the other. In whatever shape I turn this matter, and however I examine it, I can find nothing farther. (A: 649–50)

Hume says, very explicitly, what he *does* find in a case where two events are related as cause and effect: *contiguity, priority and constant conjunction*. He doesn't say, in this passage or elsewhere in the *Abstract*, what *else* one might have expected him to find, which Hume doesn't. He is more explicit on this in the body of his *A Treatise of Human Nature* (Book I, part iii), and his *An Enquiry Concerning Human Understanding*.¹ Hume's predecessors thought there were also *necessary connections* to be found in nature.² They thought that, when *c causes e*, there is something in virtue of which *c* produces, or brings about, or necessitates *e*: the cause has the power to produce the effect and the effect follows with *necessity* the cause. On the received reading of Hume's *Treatise*, this element of necessity is exactly what Hume does *not* find in causation, as it is in the objects: there is no place for necessity in nature. Once more, this reading of Hume is not unrelated to his own pronouncements. Compare his famous dictum: "Necessity is something that exists in the mind, not in objects" (T: 165). Accordingly, Hume has been typically read as "the great denier of necessary connections" (Lewis 1986f: ix).

In the last 20 or so years, however, there has been an altogether different reading of Hume's work on causation, whose origins can be found in Norman Kemp Smith's (1941) authoritative commentary on Hume's *Treatise* and in John P. Wright's (1973) work. Wright proclaimed that far from being a reductivist about causation and an eliminativist about real necessity in nature, Hume was a "sceptical realist". He was, we are told, a "causal realist" because he accepted the view that "there are real causes in nature" (1973: 127), that is, that there are objective necessary connections between events in nature. But, the claim goes on, Hume was a *sceptic* about our understanding and knowledge of them (cf. 1973: 144). This revisionary interpretative strand has been reinforced by Edward Craig (1987) and has found its *magnum opus* in Strawson (1989). These new readings of Hume have led to what Kenneth Winkler (1991) has aptly called "the New Hume". Craig goes as far as to state confidently:

Off the agenda now is the idea that [Hume] taught a strict regularity theory: that there is nothing in reality but regular sequence, and that that is accordingly all that causality amounts to, either in our concept of it or in things and events themselves. True, the tendency to speak of regularity theories as "Humean" persists, but unless it is meant . . . as nothing more than a label without historical connotations, this usage just betokens a limited acquaintance with the work of Hume.

(2000: 113)

These pronouncements might be premature. Even if it can be argued that, for Hume, what we *mean* when we talk of causation is not just regular sequence, it is not so easy to argue that for him causation, *as it is in the world*, is something more than regular sequence. Be that as it may, what I plan to do in this chapter is go through Hume's reasoning in some detail, in the hope that, in the end, we shall have a better understanding of his views on causation and their philosophical implications. In the final section, I shall engage in a discussion of the so-called "New Hume".

Before we proceed, a note on terminology is in order. Different philosophers use the term *causal realism* in different ways. Strawson (1989: 84), for instance, calls causal realism the view that "there is something about the fundamental nature of the world in

virtue of which the world is regular in its behaviour". Michael Tooley (1987: 246), on the other hand, calls causal realism the anti-reductive view that "the truth-values of causal statements are not, in general, logically determined by non-causal facts". On both characterizations, RVC would not be a causal realist position. However, it would be wrong to conclude from this that RVC is an anti-realist position. Contra Strawson's causal realism, advocates of RVC accept that it is regularities *all the way down*, and yet also accept that these regularities are *real, objective and mind-independent*. Similarly, advocates of RVC accept, contra Tooley's causal realism, that causation reduces to regularity, and yet they accept that these regularities are *real, objective and mind-independent*. So an advocate of RVC is (or can be) a realist about regularities. In so far as causation reduces to regularities, an advocate of RVC can then be a realist about causation. With these clarifications in mind, let's reserve the term "causal realism" for those views that assert that there are objective necessary connections between events in nature as well as for those views that deny that causation is reducible to non-causal facts. And let us say that, without being causal realist, RVC is a *causal objectivist* position in the sense that the regularities that causation reduces to are fully objective and mind-independent.

1.2 The two projects

Hume (T: 74) states his aim very explicitly: "This relation [of causation], therefore, we shall endeavour to explain fully before we leave the subject of the understanding." Why was he interested in the study of causation? His answer, as it is expressed succinctly in the *Abstract*, is this:

'Tis evident that all reasonings concerning *matters of fact* are founded on the relation of cause and effect, and that we can never infer the existence of one object from another, unless they be connected together, either mediately or immediately. In order therefore to understand these reasonings, we must be perfectly acquainted with the idea of a cause . . . (A: 649)

So the relation of causation underpins all our reasoning about matters of fact. Of the three "philosophical relations" that relate matters of

fact (i.e. objects in the world, or impressions), namely “identity”, “situations in time and space” and “causation”, only causation is special in the sense that it can take us “beyond what is immediately present to the senses” and can “produce a connexion” between objects that are not immediately perceived (cf. T: 73). But causation is not *just* a philosophical relation, that is, a relation that obtains between objects in the world, or impressions. It is also a “natural relation” (T: 15), that is, a relation with which the *mind* operates: it is such that it “produces an union among our ideas” (T: 94). It is because causation is a natural relation that “we are able to reason upon it, or draw any inference from it” (T: 94).³ This last observation is very important to Hume because causal reasoning seems to be somewhat the analogue of demonstrative reasoning when it comes to matters of fact. As demonstrative reasoning extends our knowledge beyond what is immediately given in *intuition* (cf. T: 70), so causal reasoning seems to extend our knowledge beyond what is immediately given in experience.

It would be wrong, however, to think that Hume’s only aim was to explain the nature of causal *reasoning*. His project has two aspects, as he thinks we can approach causation in two ways: as a “philosophical” relation and as a “natural” one. It can be argued that analysing causation as a “philosophical relation” aims to unravel what can be legitimately said of causation as it is in the objects, whereas treating it as a natural relation aims to unravel the feature of causation in virtue of which it is involved in reasoning. These two aspects of his project will lead to his two definitions of causation (see section 1.9).

1.3 Impression hunting

One major constraint of Hume’s account of causation is his empiricist epistemology. The cornerstone of this epistemology is the thought that “all our ideas, or weak perceptions, are derived from our impressions, or strong perceptions, and that we can never think of any thing we have not seen without us, or felt in our own minds” (A: 647–8; cf. also T, 4). Let’s call this the Basic Methodological Maxim. Put in a nutshell, it asserts: *no impressions in, no ideas out*. Ideas are nothing but “faint images” of impressions “in thinking and reasoning” (T: 1). This is not the place to examine Hume’s

theory of ideas. What concerns us is what concerns him *vis-à-vis* causation: what is the impression of the idea of causation? It is essential to his project to show that *there is* such an impression. For if there was not, and if the Basic Methodological Maxim was accepted, then the whole idea of causation would become vacuous (it could not exist, or in modern terms, it would be meaningless). But he does not doubt that we have this idea (cf. T: 74–5). Hume notes that this idea cannot stem from a quality (or property) of an object. Being a cause is not a particular quality of an object. It's not like being red, or being square. So to say that *c* is a cause is simply a way to describe *c* (in relation to an effect *e*) and not a way to ascribe a property to *c*. Hence, the idea of causation *cannot* derive from the impression of a property (quality) of an object. It follows that the idea of causation “must be derived from some *relation* among objects” (T: 75). What are the “essential” characteristics of this relation? They are at least two:

- spatial *contiguity* (or the presence of “chains of causes” if the two objects are not contiguous, cf. T: 75)
- temporal *succession*: “that of PRIORITY of time in the cause before the effect” (T: 76).

Hume, however, thinks that contiguity and succession are *not* sufficient for causation: they cannot “afford . . . a complete idea of causation” (T: 77). For, “an object may be contiguous and prior to another, without being consider'd as its cause” (*ibid.*). Let's call *coincidental* a sequence of events *c* and *e* such that they are spatiotemporally contiguous, *c* precedes *e*, but *c* is not the cause of *e*. And let's call *causal* a sequence of events that is not coincidental. If contiguity and succession cannot afford the basis for a distinction between a causal sequence and a coincidental one, what can? Although it is still quite early in his project, Hume is adamant in claiming that when we restrict ourselves to *particular* sequences, there is nothing beyond contiguity and succession to be discovered: “We can go no *farther* in considering this particular instance” (*ibid.*). So, when it comes to examining a particular instance (such as the collision of two billiard balls), there is nothing that can distinguish between this instance's being a causal sequence and its being merely coincidental. We “would say nothing”, Hume (T: 77) adds,

if we were to characterize a causal sequence in terms of expressions such as *c produces e*. For, the idea of “production” is synonymous with the idea of causation, and hence it would offer no further illumination.

Hume acknowledges that what is taken to distinguish between causal sequences and coincidental ones is that only the former involve some kind of necessary connection between events *c* and *e*. Hence, since contiguity and succession do not exhaust the characterization of causation, “NECESSARY CONNEXION” should also “be taken into consideration” (T: 77). So

- *necessary connection* (“and that relation is of much greater importance, than any of the other two above-mention’d” (T: 77)).

A thought that presents itself at this point is that part of the *meaning* of the idea (concept) of causation is the idea of necessary connection. Hume’s insistence on the necessary connection has led Kemp Smith to argue that Hume was far from advocating RVC. For Hume, Kemp Smith argues, “causation is more than sequence, and more also than invariable sequence. We distinguish between mere sequence and causal sequence; and what differentiates the two is that the idea of necessitation (determination or agency) enters into the latter as a quite essential element” (1941: 91–2). Be that as it may, Hume does take necessary connection to be the characteristic of causation (or, at least, the characteristic *attributed to* causation) that merits analysis. He does not have to *deny* that the idea of necessary connection is conceptually constitutive of causation. All he needs to do is *explain* the possession of this idea in a way conformable to his own epistemology. This is a pressing issue for him for the following reason. The idea of necessary connection is an idea that we do possess, but whose origin Hume is unable to find either in the “known qualities of objects” or in their “relations”. Necessity, as Kemp Smith (1941: 369) has put it, is the “essential differentia” of causation. But Hume argues that we cannot find this differentia “when we look about us towards external objects, and consider the operation of causes” (E: 63), that is, when we consider causation as a relation in the world (i.e. as a “philosophical relation”). He insists that “we are never able, in a

single instance, to discover any power or necessary connexion; any quality, which binds the effect to the cause, and renders the one an infallible consequence of the other" (E: 63). Is it, then, an idea without a corresponding impression? If so, his Basic Methodological Maxim would be refuted: we would be in possession of an idea without a corresponding impression. Hume did entertain this possibility, but considered it implausible, since, as he said, his Basic Methodological Maxim "has already been so firmly establish'd, as to admit of no farther doubt" (T: 77). Hence, Hume should try to explain how the idea of necessary connection arises; how, that is, it enters into our minds.

In a rather astonishing move, Hume abandons the route he has chosen, namely, the direct hunt for an impression that leads to the idea of necessary connection, in an attempt to ground this idea to impressions in a roundabout way. What Hume does at this juncture is shift his attention from the project of analysing causation as a "philosophical relation" – which was proved futile as an attempt to reveal the origin of the idea of necessary connection – in order to look at causation as a "natural relation".

1.4 Constant conjunction

What happens when we engage in causal inference? Hume's answer is captivatingly simple. We have memory of past co-occurrences of (types of) events *C* and *E*, where *C*s and *E*s have been directly perceived, or remembered to have been perceived. This co-occurrence is "a regular order of contiguity and succession" among tokens of *C* and tokens of *E* (T: 87). So when, in a *fresh instance*, we perceive or remember a *C*, we "infer the existence" of an *E*. Although in all past instances of co-occurrence, both *C*s and *E*s "have been perceiv'd by the senses and are remember'd", in the fresh instance, *E* is not yet perceived, but its idea is nonetheless "supply'd in conformity to our past experience . . . Without any further ceremony, we call the one [*C*] *cause* and the other [*E*] *effect*, and infer the existence of the one from that of the other" (*ibid.*). This is a basic psychological inferential procedure by which the observed past constant co-occurrence of *C*s and *E*s leads us to conclude (and to form the belief) that upon the fresh perception of a *C*, an *E* will (or *must*) follow. What is important in this process of causal inference is that it reveals "a new

relation betwixt cause and effect”, a relation that is different from *contiguity*, *succession* and *necessary connection*:

- *constant conjunction*

It is this “CONSTANT CONJUNCTION” (T: 87) that is involved in our “pronouncing” a sequence of events causal. Hume says that contiguity and succession “are not sufficient to make us pronounce any two objects to be cause and effect, unless we perceive, that these two relations are preserv’d in several instances” (*ibid.*). Remember that contiguity and succession are characteristics of single sequences. But they are not enough to exhaust the idea of causation. The “new relation” – constant conjunction – is a relation *among* sequences. It says: “like objects have always been plac’d in like relations of contiguity and succession” (T: 88). So ascriptions (“pronouncements”) of causation cannot be made of single sequences: we first need to see whether a certain sequence instantiates a constant conjunction. Does that simply mean that a sequence *is* causal (if and) only if it instantiates a constant conjunction among the relevant event types?

There is no straightforward answer to this question – at least, not yet. Hume has been describing the inferential procedure by which we move from causes to effects and has noted that this procedure is activated when (and only when) constant conjunction is present (has been observed). From this it does *not* follow that causation, as it is in the objects, amounts to constant conjunction. Hume *cannot* identify the necessary connection with the constant conjunction. For the observation of a constant conjunction generates *no* new impression in the objects perceived. Hence, Hume cannot, in a simple and straightforward manner, identify causal sequences with (instantiations of) regularities. Hume takes pains to explain why this is so (cf. T: 88; E: 75). He points out that the mere *multiplication* of sequences of tokens of *C* being followed by tokens of *E* adds no new impressions to those we have had from observing a single sequence. Observing, for instance, a single collision of two billiard balls, we have impressions of the two balls, of their collision, and of their flying apart. These are *exactly* the impressions we have no matter how many times we repeat the collision of the balls. The impressions we had from the single sequence did not include any

impression corresponding to the idea of necessary connection. But since the multiplication of instances generates no new impressions in the objects perceived, it cannot possibly add a new impression that might correspond to the idea of necessary connection. As Hume puts it: “From the mere repetition of any past impression, even to infinity, there never will arise any new original idea, such as that of necessary connexion; and the number of impressions has in this case no more effect than if we confin’d ourselves to one only” (T: 88). So, the idea of necessary connection remains recalcitrant. There was nothing in this idea that implied that only sequences that instantiate regularities (constant conjunctions) could possess necessary connections, and hence be causal. And the new relation of constant conjunction added no new impression, which could make necessary connection be a feature of the sequences that instantiate regularities. So Hume then faces a conundrum. He did unravel a “new relation betwixt cause and effect”, but it is not the one that can lead him to a straightforward impression-based explanation of the idea of necessary connection, and hence of causation.

Yet “it wou’d be folly to despair too soon” (T: 88). The reason why constant conjunction is important (even though it *cannot* directly account for the idea of necessary connection by means of an impression) is that it is the *source* of the inference we make from causes to effects. So, looking more carefully at this inference might cast some new light on what exactly is involved when we call a sequence of events causal. As Hume put it: “Perhaps ’twill appear in the end, that the necessary connexion depends on the inference, instead of the inference’s depending on the necessary connexion” (T: 88).

1.5 Causal inference

“What is the nature of that *inference* we draw from the one [event] to the other?” (T: 78). The crucial problem here is to examine whether this is a rational inference and hence whether it belongs to the realm of Reason. This is important for the following reason. Since (a) there is no direct impression corresponding to the idea of necessary connection, (b) Hume does accept (or, at least, he does not deny) that there is this idea, and (c) he has located the *source* of this idea in causal inference, the immediate objection to his view is that the *source* of the idea of necessary connection is Reason itself.

If Reason *determined* us to infer an effect from its cause, then the necessity of this inference could well be the very idea of necessary connection, whose origin Hume has been seeking. To be sure, Hume can easily dispel this objection by noting that Reason alone does not compel us to make the inference from cause to effect. What he says repeatedly is that causal inference is not a demonstrative inference. Consequently, it cannot be justified a priori, with the lights of Reason only. We can never demonstratively infer the effect from its cause, since we can always *conceive* without contradiction the effect without the cause, or the cause without the effect.⁴ Here is a typical statement of this point:

There is no object, which implies the existence of any other if we consider these objects in themselves, and never look beyond the ideas which we form of them. Such an inference wou'd amount to knowledge, and wou'd imply the absolute contradiction and impossibility of conceiving any thing different. But as all distinct ideas are separable, 'tis evident there can be no impossibility of that kind. (T: 86–7)

But the real bite of the objection we are considering is that although Reason alone might not be able to justify causal inference, Reason “aided by experience” (T: 92) (and in particular, by the experience of constant conjunction) might be able to underpin the *necessity* of causal inference.

It is precisely for this reason that Hume undertakes to prove that there simply cannot be an argument that shows that we are engaged in rational – “just” (T: 89) – inference when we form causal beliefs.⁵ In order to conclude, upon the occurrence of a *C*, that an *E* will (let alone, *must*) follow, it is not enough to plug in the premise that *C*s and *E*s have been constantly conjoined in the past. We need a stronger premise; namely, that all *C*s *are* *E*s, where the universal quantifier “all” should really range over *all* *C*s and *E*s, past, present and future. As Hume notes, this required premise would have to rely on the principle that “instances, of which we have had no experience, must resemble those, of which we have had experience, and that the course of nature continues always uniformly the same” (T: 89).

Let's call this the Principle of Uniformity of Nature (PUN). Hume's point is that experience (in the form of *past* constant

conjunctions) is not enough to guide Reason in justifying causal inference. Something more is needed to rationally underpin causal inference; namely, PUN. Or, more or less equivalently, what is needed is “a presumption of resemblance” between the observed and the unobserved. Now, PUN either can be provided by pure Reason, or else it must be grounded in experience. It cannot be provided (in the sense of being justified) by pure Reason simply because it is not demonstratively true: “we can at least conceive a change in the course of nature; which sufficiently proves that such a change is not absolutely impossible” (T: 89). Here again, what does the work is the idea that past constant conjunction is *distinct* from future constant conjunction, and hence that one can conceive the former without the latter. So PUN cannot be justified a priori by pure Reason. Nor can it be grounded in experience. Any attempt to rest PUN on experience would be circular. From the observation of *past* uniformities in nature, it cannot be inferred that nature is uniform, unless it is assumed what was supposed to be proved; namely, that nature *is* uniform – that is, that there is “a resemblance betwixt those objects, of which we have had experience [i.e. past uniformities in nature] and those, of which we have had none [i.e. future uniformities in nature]” (T: 90). In his first *Enquiry*, Hume is even more straightforward: “To endeavour, therefore the proof of this last supposition [that the future will be conformable to the past] by probable arguments, or arguments regarding existence, must evidently be going in a circle, and taking that for granted, which is the very point in question” (E: 35–6).

Hume’s conclusion is that neither Reason alone, nor Reason “aided by experience”, can justify causal inference. This has come to be known as his inductive scepticism, but it is a corollary of his attempt to show that the idea of necessary connection cannot stem from the supposed necessity that governs causal inference. For, whichever way you look at it, talk of necessity in causal inference is unfounded. Hume summed up this point as follows:

Thus not only our reason fails us in the discovery of the *ultimate connexion* of causes and effects, but even after experience has inform’d us of their *constant conjunction*, ’tis impossible for us to satisfy ourselves by our reason, why we shou’d extend that experience beyond those particular instances, which have

fallen under our observation. We suppose, but are never able to prove, that there must be a resemblance betwixt those objects, of which we have had experience, and those which lie beyond the reach of our discovery. (T: 91–2)

And elsewhere he stated:

Let men be once fully perswaded of these two principles, *That there is nothing in any object, consider'd in itself, which can afford us a reason for drawing a conclusion beyond it; and, That even after the observation of the frequent or constant conjunction of objects, we have no reason to draw any inference concerning any object beyond those of which we have had experience.* (T: 139)

Since the foregoing argument proves what Hume thought it would, namely, that causal inference does not depend “on the necessary connexion”, the road is open for his claim that “the necessary connexion depends on the inference” (T: 88). But before we see how he proceeds, it is important to examine a few serious objections to his argument so far. (Those readers who wish to carry on with Hume’s argument are advised to move directly to section 1.7.)

1.6 Necessity₁ vs necessity₂

Mackie called “necessity₁ whatever is the distinguishing feature of causal as opposed to non-causal sequences” and “necessity₂ the supposed warrant for an *a priori* inference” (1974: 12), and argued that Hume conflated these two distinct kinds of necessity. Now, Hume does seem to think that the necessity that would distinguish between causal and non-causal sequences (necessity₁) should be the necessity with which a conclusion would follow from a valid demonstrative argument (necessity₂).

In the *Enquiry* he makes this point thus:

When we look about us towards external objects, and consider the operation of causes, we are never able, in a single instance, to discover any power or necessary connexion; any quality, which binds the effect to the cause, and renders the one an

infallible consequence of the other. We only find, that the one does actually, in fact, follow the other. (E: 63)

In the *Treatise*, this point is made forcefully when he examines whether recourse to “powers” can save the idea of necessary connection. Hume considers in some detail the popular thought that what distinguishes between causal sequences and non-causal ones is that, in a causal sequence, the cause has a *power* to bring about the effect. From the fact, it might be said, that Cs have always been followed by Es in the past, we may posit that Cs have a C-power to bring about Es. So, from a fresh *c* and the concomitant C-power, we may infer that *e* will occur (cf. 1739: 90). Hume goes to some length to challenge this suggestion.⁶ But he does take it to be the case that *if* such powers existed, they would entitle us to a priori inferences from causes to effects:

If we be possest, therefore, of any idea of power in general . . . we must distinctly and particularly conceive the connexion betwixt the cause and effect, and be able to pronounce, from a simple view of the one, that it must be follow'd or preceded by the other. . . . Such a connexion wou'd amount to a demonstration, and wou'd imply the absolute impossibility for the one object not to follow, or to be conceiv'd not to follow upon the other. (T: 161–2)

Hume then insists that this “kind of connexion has been rejected in all cases” (T: 162), since there cannot be demonstrative a priori inference from cause to effect. But, in the process of his thought, he seems to use his argument against necessity₂ as a weapon against necessity₁, without offering reasons why necessity₁ is the same as necessity₂. As Mackie (1974: 13) put it, “his search for necessity₁ is sacrificed to his argument against necessity₂”. In light of this, Mackie raises three important complaints against Hume. First, Hume’s arguments against necessity₂ are correct but irrelevant, for Hume has wrongly tied causal inference to demonstratively valid inference, and has wrongly demanded that all reasons should be deductive. Secondly, although Hume did not consider “reasonable but probabilistic inferences” (1974: 15), the possibility of some kind of “non-deductively-valid argument” that is nonetheless

rational and does underpin causal inference is not “excluded by Hume’s argument” against necessity₂ (*ibid.*). So, there is still room for someone to argue against Hume that even if a cause does not necessitate its effect in the sense of necessity₂, a cause *c*, together with the constant conjunction between event-types *C* and *E*, can still “probabilify” the effect *e*. Thirdly, Hume has no argument at all against necessity₁. In the next three subsections, we shall examine these three complaints and shall try to disarm them.

1.6.1 The traditional conception of reason

Mackie claims that “Hume’s premiss that ‘reason’ would have to rely on the principle of uniformity holds only if it is assumed that reason’s performances must all be deductively valid” (1974: 15). If Mackie were right in this, then Hume’s point against necessity₂ would be valid but weak. Hume would merely point to the fact that causal inferences are not (or, cannot be) demonstrative. But if an inference need not be demonstrative to be good (or rational), Hume’s claim would be weak. Is, then, Hume’s point simply that a causal inference could never be demonstrative? By no means. Hume bases his case on a *dilemma* he poses to the traditional conception of Reason. His point is that, *by the very lights of the traditional conception of Reason*, causal inference cannot be a rational inference *either* in the sense of offering demonstrative reasons *or* in a looser sense of offering good (but not conclusive) reasons to accept the causal conclusion. This is a truly sceptical conclusion that does *not* hinge on the claim that all reasons must be deductive, a claim that Mackie falsely attributes to Hume. It amounts to the claim that the traditional conception of Reason undermines itself.

What is this traditional conception of Reason? It is the view that all beliefs should be justified (that is, backed up by *reasons*) in order to be rational. It is also the view that no inference is rational unless it is shown that it offers *reasons* to accept its conclusion. Simply put, the traditional conception of Reason craves reasons and justification. Actually, Hume’s insight is that the traditional conception of Reason is hostage to the search for a *hierarchy* of reasons, which, however, is detrimental to the rationality of causal inference that it has sought to establish. On the traditional conception of Reason, it is not enough to say that the premises of a causal inference give us

reasons to accept its conclusion. This would be an empty ritual, unless it was shown that these are, indeed, *reasons*. Causal inference itself would have to be *justified*. What this means is that one would have to offer a *further* reason *R* for the claim that the premises of the inference do give us *reasons* to rationally accept the conclusion. This further reason *R* would be a *second-order reason*: a *reason* to accept that *C*, together with the past constant conjunction of *Cs* and *Es*, is a (first-order) *reason* to form the causal conclusion that *E*. A moment's reflection shows that one would be faced with either an infinite regress or outright circularity. Hume's observation, then, is that on the traditional conception of Reason itself, causal inference remains unfounded. It cannot be justified in accordance with the demands of the traditional conception of Reason simply because the attempted justification would be question-begging.⁷

To appreciate Hume's critique of the traditional conception of Reason, let us look at the shape that the justification of causal inference could take. It can be easily seen that the following argument (AR) is *invalid*:

- | | |
|------|--|
| (AR) | |
| (PC) | All observed <i>Cs</i> have been followed by <i>Es</i> |
| (C) | A <i>C</i> is observed now |
| | |
| (E) | An <i>E</i> will occur. |

If PUN were added as an extra premise, the conclusion (E) would then logically follow from the premises. The invalid argument (AR) would be thereby turned into the valid argument (AR'):

- | | |
|-------|--|
| (AR') | |
| (PC) | All observed <i>Cs</i> have been followed by <i>Es</i> |
| (C) | A <i>C</i> is observed now |
| (PUN) | The future resembles the past |
| | |
| (E) | An <i>E</i> will occur. |

Now, Hume did argue that PUN is neither demonstratively true nor justifiable without circularity on the basis of experience. So it might be thought that his aim was *just* to show that the principle needed to turn (AR) into the deductively valid argument (AR') is

without foundation. If, indeed, this were his only aim, it would be reasonable to argue that Hume did successfully show that there was no basis to take (AR') to be a *sound* argument. But then Mackie would also be reasonable to argue that Hume's success was a Pyrrhic victory. For, surely, there is more to Reason's performances than demonstrative arguments.

Yet Hume's aim was *much* broader. Hume perceived that *on the traditional conception of Reason*, we are faced with the following dilemma. If only demonstrative inferences are taken to be rational inferences, then the so-called causal inference cannot be rational at all. For rendering a causal inference demonstrative – and hence rational – would require a *proof* of the truth of PUN, which is not forthcoming. If, on the other hand, a *looser sense* of rational inference is allowed, where we can still non-deductively infer the conclusion from the premises, provided that the premises give us good *reasons* to rationally accept the conclusion, then causal inference cannot be taken to be rational either. Why is that?

Suppose that one argued as follows. Argument (AR) above is indeed invalid, but there is no need to render it as (AR') in order to justify causal inference. For, one would continue, all we need is a *non-demonstrative*, yet *reasonable*, argument such as (AR''):

- (AR'')
- (PC) All observed Cs have been followed by Es
- (C) A C is observed now
- (R) (PC) and (C) are reasons to believe that E will occur
-
- (E) Probably, an E will occur.

Hume's general point is precisely that, by the very lights of the traditional conception of Reason, principle (R) cannot be a good reason for the conclusion (E). Not because (R) is not a deductively sufficient reason, but because any defence of (R) would be question-begging.⁸ To say, as (R) in effect does, that a *past* constant conjunction between Cs and Es is reason enough to make the belief in their *future* constant conjunction *reasonable* is just to assume what needs to be defended by further reason and argument.

1.6.2 The AP property

Mackie's second complaint against Hume is precisely that there *are* good arguments to defend the view that a *past* constant conjunction between *Cs* and *Es* is reason enough to make the belief in their *future* constant conjunction *reasonable*, provided that we allow for reasonable *probabilistic* inferences. Let's, following Strawson (1989: 111), call "AP property" a property such that if it could be detected in a causal sequence, it would bring with it "the possibility of making *a priori certain causal inferences*". Hume has conclusively shown, Mackie says, that causal sequences do not have the AP property. Yet, he adds (1974: 17), Hume has *failed* to show that they also lack the property of licensing "probabilistic *a priori* causal inferences".⁹ This is supposed to be a weakened version of the AP property: the past constant conjunction between *Cs* and *Es* makes it *a priori* more likely (although not certain) that future tokens of *C* will be accompanied by future tokens of *E*, than not. If Mackie is right, Hume has neglected a real possibility of understanding the rational nature of causal inference, and hence the nature of necessary connection. Is there anything that can be said in Hume's defence?

The reason why Hume didn't take seriously the possibility canvassed by Mackie is that to call some sequences of events *necessary* in the weaker sense that the occurrence of the first event makes *a priori more likely* the occurrence of the other would constitute no progress at all. For, one can perceive nothing *in* the sequence of *c* and *e* that points to the occurrence of *e* after *c* has been perceived. Hence one perceives nothing that can give rise to an impression that *e* is likely to occur. Hume does acknowledge that the more frequently we observe *Cs* being followed by *Es*, the more "fortified" and "confirmed" is the belief that, upon a fresh *c*, an *e* will follow (cf. E: 58). But he claims that this is the product of custom, and in particular of the "custom to transfer the past to the future".

... where the past has been entirely regular and uniform, we expect the event with the greatest assurance, and leave no room for any contrary supposition. But where different effects have been found to follow from causes, which are to *appearance* exactly similar, all these various effects must occur to the

mind in transferring the past to the future, and enter into our consideration, when we determine the probability of the event.
(E: 58)

For Hume there is simply *no* reason to think that when, given past experience, we believe that a certain future event is more *likely* to happen than not, we avoid reliance on the problematic “supposition, *that the future resembles the past*” (T: 134).

Now, Mackie thinks that Hume was wrong in this. He insists that “a probabilistic inference would not need to invoke the uniformity principle which produces the circularity that Hume has exposed” (1974: 15). To say that, given the past co-occurrence of *Cs* and *Es*, when an event of type *C* occurs an event of type *E* is likely to occur is *not* to say that the event of type *E* *will* occur; hence, it does *not* amount to asserting that the future *will* (or does) resemble the past.

Even if we were to grant to Mackie that PUN is not presupposed by a priori probabilistic inference (something which is very doubtful), it is clear that some substantive principles – which cannot, therefore, be seen as a priori true – have to be invoked. Here is an interesting example. In order to defend his view that, contra Hume, there is space for *rational* a priori probabilistic inference, Mackie (1974: 15) appeals to John Maynard Keynes’s (1921) system of inductive logic. This is not the place to review the many problems that this (as well as Carnap’s 1950) so-called logical theory of probability faces.¹⁰ It suffices for our purposes to note that Keynes (1921: 287) founded his system of inductive logic on what he called the Principle of Limited Variety (PLV). Without getting into technicalities, let’s envisage the following possibility. Suppose that although *C* has been invariably associated with *E* in the past, there is an unlimited variety of properties E_1, \dots, E_n such that it is logically possible that future occurrences of *C* will be accompanied by any of the E_i s ($i = 1, \dots, n$), instead of *E*. Then, and if we let *n* (the variety index) tend to infinity, we cannot even start to say how likely it is that *E* will occur given *C*, and the past association of *Cs* with *Es*.¹¹ So we cannot engage in probable causal inference, unless PLV, that is, a principle that excludes the possibility just envisaged, is true. Hume’s argument against PUN can now be recast against PLV. The latter is not an a priori truth; nor can it be grounded on experience, without begging the question. To call this principle synthetic a priori is of no help. It would simply beg the

question against Hume.¹² So neither Hume nor we should try to save some notion of necessity by introducing a probabilistic analogue of necessity₂.

1.6.3 Hidden powers

Mackie's (1974: 13) final objection is that Hume "has no case at all" against necessity₁; namely, whatever is the distinguishing feature of causal as opposed to non-causal sequences. This, Mackie (1974: 20) says, is supposed to be "an intrinsic feature of each individual causal sequence". But Hume did offer a rather important argument against necessity₁. He insisted that we could never observe such an intrinsic feature *in* the sequence. This, of course, does not imply that there is not such a feature. Yet Hume accepted what might be called the *Manifestation Thesis*: there cannot be unmanifestable "powers", that is, powers that exist, even though there are no impressions of their manifestations. Hence, he stresses: "'tis impossible we can have any idea of power or efficacy, unless some instances can be produc'd, wherein this power *is perceiv'd* to exert itself" (T: 160). And repeats: "If we be possest, therefore, of any idea of power in general, we must also be able to conceive some particular species of it" (T: 161). And again: "The distinction, which we often make betwixt *power* and the *exercise* of it, is equally without foundation" (T: 171). No such exercise of power is ever perceived in sequences of events, or in events that occur in our minds. But then, contra Mackie, no necessity₁ – *qua* an intrinsic feature that makes a sequence causal – is ever perceived. The Manifestation Thesis is, then, strong enough to disallow that there are such things as powers, and hence such a thing as necessity₁.

Indeed, Hume spends quite some time trying to dismiss the view that we can meaningfully talk of powers. His *first* move is that an appeal to "powers" in order to understand the idea of necessary connection would be no good because terms such as "*efficacy, force, energy, necessity, connexion, and productive quality*", are all nearly synonymous" (T: 157). Hence, an appeal to "powers" would offer no genuine explanation of necessary connection. His *second* move is to look at the theories of his opponents: Locke, Descartes, Malebranche and others. The main theme of his reaction is that all these theories have failed to show that there are such things as

“powers” or “productive forces”. As he puts it: “This defiance we are oblig’d frequently to make use of, as being almost the only means of proving a negative in philosophy” (T: 159). In the end, however, Hume’s best argument is a reiteration of his Basic Methodological Maxim: that we “never have any impression, that contains any power or efficacy. We never therefore have any idea of power” (T: 161). In so far as we take the concept of “necessary connexion betwixt objects” to mean that this connection “depends upon an efficacy or energy, with which any of these objects are endow’d”, then this concept has no “distinct meaning” (T: 162).¹³

It may, of course, be objected that Hume’s argument against necessity₁ is wrong because it is based on an excessive verificationism, as this may be evinced by Hume’s Manifestation Thesis. This may well be so. Empiricists have protested against *unmanifestable* powers, based on the epistemic argument that if they are unmanifestable, then they cannot be known to exist. From this, it does not of course follow that they *don’t* exist. This move from what can (or cannot) be known to exist to what does (or does not) exist has been dubbed the *epistemic fallacy*. And it is, indeed, a fallacy. But, in defence of Hume, it might be said that his Manifestation Thesis is more of an application of Ockham’s razor than the product of the epistemic fallacy. Ockham’s razor says: do not multiply entities beyond necessity. For Hume, positing such unmanifestable powers would be a gratuitous multiplication of entities, especially in light of the fact that Hume thinks he can explain the origin of our idea of necessity without any appeal to powers and the like. Still, Hume might be wrong in his argument. But Mackie is certainly wrong to think that Hume has no case at all against necessity₁.

1.7 Union in the imagination

Let us briefly review where we stand. I have just said that Hume denies necessity₁; namely, an *intrinsic* feature that is possessed by causal sequences and is lacked by non-causal ones. He does, however, think (or, at least, he does not deny) that the idea of necessity is part of the idea of causation. But, so far, there has been nothing to prevent this idea from being an idea of an *intrinsic* feature of a single sequence that renders it causal. If Hume had managed to identify his “new relation” of constant conjunction with necessity,

he could at least argue that necessity is an *extrinsic* feature of some sequences in virtue of which they are causal; namely, the feature that relates to the fact that some sequences instantiate regularities. But, as we saw in section 1.4, he cannot afford this identification. So he is still pressed to show what the origin of the idea of necessary connection is. Having deviated from his original impression hunting, he has endeavoured to show that we are *not* engaged in rational inference when we form causal opinion. But then, what is the foundation of causal opinion? Hume's hope is that the search for this foundation will help him pin down the mysterious idea of necessity.

Hume claims that it is the "union of ideas" by principles of the imagination, and not of Reason, that explains the formation of our causal beliefs and opinions. He introduces three basic "principles of union among ideas" in the imagination. They are: *resemblance*, *contiguity* and *causation* (T: 93). These principles are the backbone of Hume's alternative (psychological) theory of reasoning. They are "neither the *infallible* nor the *sole* causes of an union among ideas" (T: 92). However, they are "the only *general* principles, which associate ideas" (T: 92–3). The positive argument for his theory of belief formation is "Had ideas no more union in the fancy than objects seem to have to the understanding, we coul'd never draw an inference from causes to effects, nor repose belief in any matter of fact. The inference, therefore, depends solely on the union of ideas" (T: 92). The argument is a *reductio ad absurdum* of the view that Reason governs causal inference and belief formation. If causal inference and belief formation had the features that the traditional conception of Reason demanded of them, then there would be no transition from cause to effect, nor belief formation. But there are both. Hence, the way in which the traditional conception of Reason conceives causal inference and belief formation is wrong. Causal inference and causal belief formation are governed by different principles. What makes, on Hume's account, causal inference (or, better, causal transitions from *C* to *E*), and belief formation, possible is that the principles of the imagination are in operation.

All this may lead us better to understand Hume's claim that causation is a philosophical relation as well as a "natural relation" (T: 94). Although causation is subject to philosophical analysis (whatever the consequences of this might be), it is also a "natural relation" in that it is a (the?) way in which the mind operates. This

is an empirical claim, but for Hume it is an important one because “’tis only so far as [causation] is a *natural* relation, and produces an union among our ideas, that we are able to reason upon it, or draw any inference from it” (T: 94). In other words, causation might be the object of our analysis, but it is also *presupposed* (as an empirical principle of human psychology) for the functioning of the mind. What Hume, in effect, does when he describes the mechanics of causal inference is to leave aside his analysis of causation as a “philosophical relation” and to concentrate on its role in inference *qua* a natural relation. His hope is that by looking at causation as a natural relation, he will discover the missing element of his analysis of causation as a philosophical relation, that is, the origin of the idea of necessary connection.

There are two principles of the imagination that are needed for the explanation of the formation of causal beliefs: first, the principle that an observed constant conjunction creates a “union in the imagination” between tokens of two event-types; secondly, the principle that a present impression transmits some of its force or vivacity to an associated idea. Hume puts it like this:

We have no other notion of cause and effect, but that of certain objects, which have been *always conjoin’d* together, and which in all past instances have been found inseparable. We cannot penetrate into the reason of the conjunction. We only observe the thing itself, and always find that from the constant conjunction the objects acquire an union in the imagination. When the impression of one becomes present to us, we immediately form an idea of its usual attendant; and consequently we may establish this as one part of the definition of an opinion or belief, that *’tis an idea related to or associated with a present impression*. (T: 93)

Hume swiftly moves from the “idea of its usual attendant” to the “belief” that the usual attendant will occur. This is not surprising, since for him, a belief is just a different “manner” in which we conceive an object. A belief “can only bestow on our ideas an additional force or vivacity” (T: 96). But what really matters to his argument is that a causal belief “arises immediately, without any new operation of the reason or the imagination” (T: 102). Take, for instance, the

belief that an event *e* will happen after there appears in the mind an impression or idea of an event *c*, and perceptions or memories of the constant conjunction between *Cs* and *Es*. On Hume's theory, what we would otherwise analyse as an inference (or a movement of thought) is no inference at all, since it happens automatically and unconsciously. There is no movement of thought, as it were. The belief that *e* will happen is as "unavoidable as to feel the passion of love, when we receive benefits; or hatred, when we meet with injuries" (E: 46). What governs this "immediate" transition is "Custom or Habit" (E: 43). So, from the point of view of the *description* of what happens in the mind, Hume's point is that what has appeared to be an inference is nothing but a "customary transition" (T: 103) from a certain impression or memory of an object to a lively idea of its usual attendant, where the whole process is *conditioned* by the observations of past co-occurrences. Hume stresses repeatedly that this operation of the mind is "immediate": "the custom operates before we have time for reflection" (T: 104). In particular, the mind does not rely on PUN in order to draw causal conclusions. In a certain sense, if it did, it would never draw any conclusions. As he puts it: "... the understanding or the imagination can draw inferences from past experiences, without reflecting on it [the Principle of Uniformity of Nature]; much more without forming any principle concerning it, or reasoning upon that principle" (*ibid.*).

What exactly is this custom, of which Hume says that it is "the great guide of human life" (E: 44)? Without going into much detail, we should note that Hume takes custom to be a central posit of his own psychological theory of belief formation. It is "a principle of human nature" (E: 43). And although he stresses that this principle is "universally acknowledged", what matters is that for him it is "the ultimate principle, which we can assign, of all of our conclusions from experience" (*ibid.*). So Hume refrains from explaining further this "ultimate principle". He just posits it, as a "cause" whose own "cause" he does not "pretend to give" (*ibid.*). As we shall see below, it is not accidental that Hume characterizes custom as a "cause". It is custom that, according to Hume, causes us to draw causal conclusions and to form causal beliefs. Hume's retreat to *causal discourse* in his attempt to unravel the foundation of causal inference – the "customary transition" – will prove indispensable for the completion of his search for the origin of the idea of necessary connection.

1.8 Necessary connection

Hume now feels fully equipped to reveal to us how, in the end, “the necessary connexion depends on the inference” (T: 88). But it’s not the rationalists’ inference that holds the key to the idea of necessity. It is the “customary transition” that he has put in its place. His summary of his long argument is very instructive:

Before we are reconcil’d to this doctrine, how often must we repeat to ourselves, *that* the simple view of any two objects or actions, however related, can never give us any idea of power, or of a connexion betwixt them: *that* this idea arises from the repetition of their union: *that* the repetition neither discovers nor causes any thing in the objects, but has an influence only on the mind, by that customary transition it produces: *that* this customary transition is, therefore, the same with the power and necessity; which are consequently qualities of perceptions, not of objects, and are internally felt by the soul, and not perceiv’d externally in bodies? (T: 166)

So let us pick up the threads. The only new element that has entered Hume’s analysis of causation as a philosophical relation (i.e. as a relation in the world) is constant conjunction. The idea of constant conjunction – which is *also* the driving force behind the customary transition from cause to effect – does not arise from any new impression in the objects, yet it is the *source* of the further idea of necessary connection. How is this? As he notes, “after frequent repetition I find, that upon the appearance of one of the objects, the mind is *determin’d* by custom to consider its usual attendant, and to consider it in a stronger light upon account of its relation to the first object” (T: 156). He immediately adds, “’Tis this impression, then, or *determination*, which affords me the idea of necessity”.

This is slightly odd. It seems that Hume just *posits* a new impression, “determination”, which will carry the weight of his explanation of the origin of the idea of necessary connection. Hume indeed starts with an aspect of his own positive theory, namely, that habit or custom operates on the mind to make it form a *belief* of the usual attendant of an object, and takes this *aspect* of his theory as a *datum* that will give rise to the required impression. It’s not accidental then that Hume appended his analysis of causal inference with an

exposition of his own psychological theory of causal belief formation – which we saw in the previous section. For it turns out that his own psychological theory is essential for the completion of his original task; namely, his impression hunting. There is *something* that occurs in the mind as a result of the observation of constant conjunction. This something is not an “impression of sensation”. If it were, the observation of a single instance would have the same effect on the mind. But it does not. This something, as Stroud (1978: 43) has nicely put it, is “a peculiar feeling that arises from the repeated occurrence of associated perceptions”. Hume calls it an “internal impression, or impression of reflection” (T: 165). In the *Enquiry*, he calls it a “sentiment” (E: 75). But all this does not matter much, I think. That there must be an impression corresponding to the idea of necessary connection follows from Hume’s Basic Methodological Maxim. That it isn’t an “impression of sensation” follows from his analysis of what is perceived *in* the objects. That, nonetheless, something happens to the mind when a “multiplicity of resembling instances” is observed follows from his own positive psychological theory of causal belief formation, that is, from his own account of causation as a “natural relation”. Then, it *must* be the case that this something that happens to the mind is the sought-after impression. This something that happens to the mind is what Hume calls the feeling of “determination”. Indeed, Hume notes: “this determination is the only effect of the resemblance; and therefore must be the same with power or efficacy, whose idea is derived from the resemblance” (T: 165). Its presence in the human mind after the observation of “resemblance in a sufficient number of instances” (T: 165) is, as Stroud (1977: 86) has rightly put it, “simply a fundamental fact about human beings that Hume does not try to explain”.

It’s not accidental that Hume retreats to *causal talk* to state this fundamental fact about human beings.¹⁴ He stresses that the idea of power or connection “is copy’d from some effects of the multiplicity, and will be perfectly understood by understanding these effects” (T: 163). And since the “multiplicity of resembling instances” has no effects on the objects involved in it, the effects we look for should be effects “in the mind” (T: 165). Ultimately, “Necessity is then the *effect* of this observation [of constant conjunction], and is nothing but an internal impression of the mind, or a determination to carry our

thoughts from one object to another" (T: 165, emphasis added). Note that Hume talks of the "effect" of constant conjunction in the mind. Note also that the very concept of determination is itself causal. His theory of how the idea of necessity arises is, then, a *causal* theory. This is really important. For Hume's central posit (the feeling of determination) is part of his causal theory of the origin of the idea of necessity. As Stroud (1977: 92) stresses, Hume offers "a causal explanation of how and why we come to think of things in our experience as causally connected".¹⁵

Having thus arrived at the sought-after "impression", Hume can come back to his "suspicion" that "the necessary connexion depends on the inference, instead of the inference's depending on the necessary connexion" (T: 88) in order to substantiate it. These two "are, therefore, the same", Hume says (T: 165). For "The necessary connection betwixt causes and effects is the foundation of our inference from one to the other. The foundation of our inference is the transition arising from accustom'd union" (*ibid.*). Note that for Hume this customary transition – the "foundation of our inference" – is not something within our control.¹⁶ Our minds just have this propensity to perform these customary transitions as the result of being *determined* to do so by the observation of "resemblance in a sufficient number of instances" (*ibid.*).

Where does all this leave the idea of necessary connection? Hume has finally unpacked the "essence of necessity" and has found that it "is something that exists in the mind, not in objects" (*ibid.*). Power and necessity "are consequently qualities of perceptions, not of objects, and are internally felt by the soul, and not perceiv'd externally in bodies" (T: 166). In his first *Enquiry*, Hume sums up his point thus:

The first time a man saw the communication of motion by impulse, as by the shock of two billiard-balls, he would not pronounce that the one event was *connected*; but only that it was *conjoined* with the other. After he has observed several instances of this nature, he then pronounces them *connected*. What alteration has happened to give rise to this new idea of *connexion*? Nothing but that he now *feels* these events to be *connected* in his imagination, and can readily foretell the existence of one from the appearance of the other. When we say,

therefore, that one object is connected with another, we mean only, that they have acquired a connexion in thought, and give rise to this inference, by which they become proofs of each other's existence. (E: 75–6)

But, surely, when we ascribe necessity to a sequence of events, we don't ascribe something to minds that perceive them. Nor does Hume claim that we do this: "... we suppose necessity and power to lie in the objects themselves, not in our mind, that considers them" (T: 167) He does, however, claim that this supposition is *false*. Indeed, he goes on to explain *why* this false belief is widespread (even inevitable). So he claims that the idea of objective necessity is *spread* by mind onto the world: "'Tis is a common observation, that the mind has a great propensity to spread itself on external objects, and to conjoin with them any internal impressions, which they occasion, and which always make their appearance at the same time as these objects discover themselves to the senses" (*ibid.*).

He adds that the "propensity" of the mind to spread itself onto the world is the "reason" why we suppose that there is necessity "in the objects we consider" (*ibid.*). In the *Enquiry*, he makes a similar point when he notes that "as we *feel* a customary connexion between the ideas, we transfer that feeling to the objects; as nothing is more usual than to apply to external bodies every internal sensation, which they occasion" (E: 78).

How, then, are we to understand Hume's position? Hume disavows the view that causation is, somehow, mind-dependent. As he (T: 168) stresses, he "allows" that "the operations of nature are independent of our thought and reasoning", as he also allows that "objects bear to each other the relations of contiguity and succession; that like objects may be observ'd in several instances to have like relations; and that all this is independent of, and antecedent to the operation of understanding" (*ibid.*).

What he does *not* allow is our going "any farther" in order to "ascribe a power or necessary connexion to these objects" (T: 169) So one might say that for Hume there is causation in the world, but if there is anything like an objective content to the talk about necessary connections in the objects, then this is exhausted by the regularities (constant conjunctions) of which they partake. And if we

think that there is any extra content in the talk of necessary connections *in* the objects, “this is what we cannot observe in them, but must draw the idea of it from what we feel internally in contemplating them” (T: 169). Hume, then, can be seen as offering an objective theory of causation in the world, which is however accompanied by a mind-dependent view of necessity. This dual aspect of Hume’s account of causation is reflected in his two definitions of causation, to which we shall now turn. After we have examined them, we shall come back to the issue of what (if any) is the right interpretation of Hume’s views.

1.9 Two definitions of “cause”

At the very end of his enquiry into causation, Hume suggests that it is now time to offer “an exact definition of the relation of cause and effect”, which is supposed to “fix [the] meaning” of this relation (T: 169). But he goes on to offer *two* definitions that “are only different, by their presenting a different view of the same object” (T: 170). The first definition (Df_1) is “We may define a CAUSE to be ‘An object precedent and contiguous to another, and where all the objects resembling the former are plac’d in like relations of precedency and contiguity to those objects, that resemble the latter’” (*ibid.*). The second definition (Df_2) runs: “A CAUSE is an object precedent and contiguous to another, and so united with it, that the idea of the one determines the mind to form the idea of the other, and the impression of the one to form a more lively idea of the other” (*ibid.*).

Surprisingly, Hume offers no further defence (or analysis) of these two definitions. He takes it that his arguments so far lead naturally to them. He does claim that the first definition is a definition of causation as a “philosophical relation”, while the second is a definition of causation as a “natural relation”. But, as we have already noted, he thinks that they both define the same relation. What is more interesting is that he thinks that *both* definitions entail that “there is no absolute, nor metaphysical necessity” (T: 172).

In the first *Enquiry*, where he restates the two definitions (with some interesting alterations),¹⁷ he stresses that both definitions are “drawn from circumstances foreign to the cause” (E: 77). A “perfect” definition would point to “that circumstance in the cause, which gives it a connexion with its effect” (*ibid.*). But Hume makes

clear once more that “we have no idea of this connexion; nor even any distinct notion what it is we desire to know, when we endeavour at a conception of it” (*ibid.*). All he can therefore do is illustrate how the definitions are supposed to work by offering a typical example of a causal claim such that “the vibration of this string is the cause of this particular sound”. This, it should be noted, is a claim that has the surface structure of a singular causal proposition. Although it seems to imply that what makes this sequence of events causal is something intrinsic to the sequence, Hume is adamant that what we mean by that “affirmation” is either that “*this vibration is followed by this sound, and that all similar vibrations have been followed by similar sounds*” or that “*this vibration is followed by this sound, and that upon the appearance of one, the mind anticipates the senses, and forms immediately an idea of the other*” (*ibid.*).

Hume’s two definitions – and especially his puzzling remark that they are both drawn from terms “foreign” to causation – have generated quite an impressive interpretative literature that cannot be properly discussed here.¹⁸ It seems to me, however, that we can see why (a) he has to offer two definitions; and (b) he thinks that they are both drawn from factors “foreign” to the cause. As I have already noted, Hume has aimed at a dual target. On the one hand, he has aimed to analyse causation as it is in the objects (that is, as a philosophical relation); on the other hand, he has been led to consider causation as a natural relation; namely, as a principle with which the mind operates. There is no reason to think that this is not one and the same relation. But Hume’s analysis of causation as a natural relation has found in it the elements that he couldn’t find in his analysis of causation as a philosophical relation. In particular, although he couldn’t find in causation, as it is in the objects, anything that could correspond to the idea of necessary connection, he found the corresponding impression in his account of causation as a principle with which the mind operates.

If indeed the concept of causation he ventured to explain is one single concept, and if, in particular, its full grasp would offer “that circumstance in the cause, which gives it a connexion with its effect” (E: 77), it should be clear that Hume’s dual project, namely, looking at causation as a philosophical relation and as a natural one, has failed to offer a unified account (definition) of the concept of causation. Hence, all Hume can do is offer the results of his dual

investigation – in the form of the two definitions above – and proclaim that although they are both aimed at one and the same concept, each of them offers only *some* aspect of this concept. If Df_1 was all there was to the concept of causation, then causation would have nothing to do with necessity. Construed as an exclusive definition, Df_1 is a typical version of RVC. In effect, it states that causation amounts to invariable succession. But Hume has spent most of his time trying to unravel the origin of the mysterious idea of necessary connection. So he feels that it may be objected to Df_1 that it is “drawn from objects foreign to the cause” (T: 170). For, there is no mention of necessity in it. He cannot say anything more about causation *as it is in the world*. In fact, he says that nothing more can be said. All he can then do is draw attention to another aspect of causation, as this is captured by Df_2 . This does make reference to a concept of necessitation (“determination”), but it also introduces *minds* into the definition of causation. It makes a condition of an event causing another event that “a mind observes and reacts to what it observes” (Robison 1977: 160). Hume is bound to feel that Df_2 , taken in isolation, is also drawn from elements foreign to the cause, not least because it *seems* to compromise the mind-independent character of causation – an aspect that was brought to light by Df_1 . But he disavows any attempt to find “a juster definition” in their place. He just repeats briefly the line of reasoning that led him to the two definitions (cf. T: 170).

What needs to be stressed is that according to *both* definitions, an individual sequence of events is deemed causal only because something *extrinsic* to the sequence occurs, be it the constant conjunction of similar events, as in Df_1 , or the customary transition of the mind from the appearance of the one, to the idea of the other, as in Df_2 . But there is an interesting way in which each definition supplements the other. The extrinsic feature of a sequence that makes it causal according to Df_1 (i.e. the instantiation of a regularity) *is* the feature that conditions the mind to think of this sequence as necessary. And the extrinsic feature of a sequence that makes it causal according to Df_2 (i.e. the felt determination of the mind) *is* the feature of the mind that *responds* to some objective condition in the world. As Stroud (1977: 89) has pointed out, if nothing fulfilled the conditions of Df_1 , that is, if there were no regularities in nature, our minds would not form the idea of causation – and especially,

the problematic idea of necessary connection. Conversely, it is because there are minds – which are such that they fulfil the conditions of Df_2 – that “any things in the world are thought to be related *causally* or *necessarily* at all” (Stroud 1977: 90).¹⁹

What seems to me quite striking is that, in a rather astonishing – and relatively unnoticed – passage of the first *Enquiry*, and long before he offered the two definitions, Hume made an intriguing suggestion as to *why* there is a coincidence between the conditions under which the two definitions hold. Nature, Hume says, did not leave it up to us to draw the right causal conclusions and to form the right causal beliefs, but made sure that there is

a pre-established harmony between the course of nature and the succession of our ideas; and though the powers and forces, by which the former is governed, be wholly unknown to us; yet our thoughts and conceptions have still, we find, gone on in the same train with the other works of nature. Custom is that principle, by which this correspondence has been effected; so necessary to the subsistence of our species, and the regulation of our conduct, in every circumstance and occurrence of human life. . . . As nature has taught us the use of our limbs, without giving us the knowledge of the muscles and the nerves, by which they are actuated; so she has implanted in us an instinct, which carries forward the thought in a correspondent course to that which she has established among external objects; though we are ignorant of those powers and forces, on which this regular course and succession of objects totally depends.

(E: 54–5)

What matters here is Hume’s claim that, in effect, causation as it is in the objects (“the course of nature”) and causation as we take it to be (“succession of our ideas”) are in “harmony”. In light of his subsequent discussion of the two definitions, Hume seems to suggest that nature has made it the case that Df_1 (causation as the course of nature) and Df_2 (causation as succession of ideas) go hand-in-hand. This does not mean that we don’t make mistakes in calling a sequence causal. The principles of the imagination are far from “infallible”. But it does mean that Df_1 and Df_2 cannot be such that they are systematically out of step with each other.

I think it is important to stress that Hume ended up with a *double-aspect* view of causation, which reflected his dual aim.²⁰ In this light, Stroud is right in stressing that Df_1 expresses “all the objective relations that actually hold between events we regard as being causally related” (1977: 91), whereas Df_2 expresses the *extra content* of our belief in causation – where this extra content is something that has its origin in the mind and is (*falsely*) projected onto the world.

Did Hume, then, endorse the RVC? I think there are reasons to go against the prevailing tide and argue that, to some extent at least, he did. To the extent to which we can have an account of causation as *it is in the objects*, causation can only be invariable succession. But this is not to say that what we *mean* when we talk about causation is captured by RVC. Nor did Hume think that we do. I take it, however, that one of his major contributions was to make possible an *error theory* about objective necessity: there is a belief that necessity and the cognate are objective qualities in the objects, but this belief is *false*. Do we then have to *reform* the concept of causation so that its new meaning is fully given by RVC and contains no reference to necessity? Do we have to trim down the content of the concept of causation so that it is equated with RVC? I think this is an open question. But I also feel that Hume would find this task impossible. Being a “natural relation”, causation is so ingrained into our lives and modes of thinking (and so usefully so), that it would be a hopeless (and maybe pointless) task to embark on a reform of its meaning.

1.10 A new Hume?

The advocates of the “New Hume” claim that any attempt to view Hume as espousing RVC is fundamentally wrong. Indeed, there are certain passages of the *Treatise*, and especially of the first *Enquiry*, that suggest that Hume did allow that there is something, an “ultimate connection” (T: 91), in virtue of which a regularity holds, although we shall never be able to comprehend what this is. Consider some of them:

... we are ignorant of those powers and forces, on which [the] regular course and succession of objects totally depends.

(E: 55)

It must certainly be allowed, that nature has kept us at a great distance from all her secrets, and has afforded us only the knowledge of a few superficial qualities of objects; while she conceals from us those powers and principles on which the influence of these objects entirely depends. (E: 32–3)

But notwithstanding this ignorance of natural powers and principles, we always presume, when we see like sensible qualities, that they have like secret powers, and expect that effects, similar to those which we have experienced, will follow from them. (E: 33)

The scenes of the universe are continually shifting, and one object follows another in an uninterrupted succession; but the power or force which actuates the whole machine, is entirely concealed from us, and never discovers itself in the sensible qualities of body. (E: 63–4)

. . . experience only teaches us, how one event constantly follows another; without instructing us in the secret connexion, which binds them together, and renders them inseparable. (E: 66)

For some, such as Mackie (1974: 21), Hume “may well have his tongue in his cheek” when he makes assertions such as the above. But Strawson thinks that he did not. Strawson calls “Causation” (with a capital C) a view of causation that is substantially distinct from RVC. To believe in Causation is to believe “(A) that there is something about the fundamental nature of the world in virtue of which the world is regular in its behaviour; and (B) that that something is what causation is, or rather it is at least an essential part of what causation is” (Strawson 1989: 84–5). Given that most present-day Humeans deny both (A) and (B), we can call Strawson’s view non-Humean. Yet his startling suggestion is that Hume too was non-Humean. On Strawson’s reading of Hume, Hume “never seriously questions the idea that there is Causation – something about the nature of reality in virtue of which it is regular in the way that it is – although he is passionate and brilliant in his attack on the view that we can know anything about its nature” (1989: 204).

Michael J. Costa has introduced a useful distinction between *causal objectivism* and *power realism*. The former is the view that “causes are objective in the sense that causal relations would continue to hold among events in the world even if there were no minds to perceive them” (1989: 173). The latter is the view that “objects stand in causal relations because of the respective causal powers in the objects” (*ibid.*). Now, causal objectivism is perfectly consistent with – if it is not directly implied by – RVC. And there has also been ample reason to think that Hume was a causal objectivist. So the question is whether Hume was a power realist.

There is an important hurdle that the power-realist interpretation of Hume should jump: his theory of ideas. It is one thing to argue – quite plausibly – that Hume’s theory of ideas (and in particular his Basic Methodological Maxim) is wrong.²¹ It is quite another thing to argue that Hume did not, really, endorse it. Craig (1987: 91) does make a first move towards this bold claim: Hume’s “theory of belief is more important than his theory of ideas”. But even if this is so, that is, even if Hume was primarily concerned with the origin of our natural belief in causation, he didn’t seem to leave behind his theory of ideas. So the next step taken by the advocates of New Hume is to stress a distinction that Hume seemed to have drawn between *supposing* that something is the case and *conceiving* that it is the case.

Both Craig (1987: 123–4) and Strawson (1989: 49–58) argue as follows. Hume’s official theory of ideas relates to what can be properly conceived. Anything that can be properly conceived, that is, anything of which we have a contentful idea, must be based on impressions. It should be, in fact, copied from impressions. But, they argue, Hume also has a theory of “supposition”, which is distinct from his theory of ideas. We can suppose that something is the case and form an intelligible view of it, even if this “supposition” is not “contentful (or intelligible) on the terms of the theory of ideas” (Strawson 1989: 54). Strawson, in particular, claims that “the special limited theory-of-ideas-based notion of ‘intelligibility’ is essentially supplemented by the notion of what is intelligible in the sense of being coherently supposable” (1989: 58). Both Craig and Strawson go on to claim that Hume did allow that we can “suppose” the existence of powers, or of necessity, where this is a “genuine supposition” (Strawson 1989: 45), and far from being

“senseless” (Craig 1987: 124). Strawson claims that Hume’s theory of “supposition” allows him to form a “relative” idea “of true causal power or force in nature”. This “relative” idea gets its content from the following description: “whatever it is in reality which is that in virtue of which reality is regular in the way that it is” (1989: 52). What is more interesting is that, according to Strawson, the above description allows us (and Hume, in particular) to “refer to [causal power] while having absolutely no sort of positive conception of its nature on the term of the theory of ideas” (*ibid.*).²² So, both Craig and Strawson conclude, Hume was a causal realist. He was a sceptic about our having a positive conception of powers in nature in so far as this scepticism sprung from his theory of ideas. Nonetheless, he believed in the existence of these powers and he propounded that we can coherently *suppose* that they exist.

This summary of the Craig–Strawson interpretation has been very brief. Their account deserves much more attention than I have given it. Its main philosophical suggestion, that one *can* be a non-Humean causal realist in the strong sense of admitting the existence of something in virtue of which regularities hold, is well taken. As we shall see in some detail in Chapter 6, there is indeed a non-Humean philosophical school, which tries to explain why regularities hold by positing the existence of relations of necessitation between properties. But the present question is whether *Hume* was a non-Humean causal realist. Did he really take the view that Craig and Strawson attribute to him?

This view has been challenged in two ways. The first is by Winkler (1991). He has offered a very careful study of all the passages that might suggest a realist interpretation of Hume, and has – I think, persuasively – argued that the hurdle that the causal realist interpretation of Hume should jump, namely, his theory of ideas, is insurmountable. In particular, he claims that the distinction between “supposing” and “conceiving” is not well founded: “an interest in *acts of supposing* or *relative ideas* is no sign that we have moved into territory where the theory of ideas does not hold sway” (1991: 556). For Winkler, in order to preserve Hume’s scepticism it is enough to suppose that he refused to *affirm* the existence of something other than regularity in nature, “a refusal rooted in the belief that there is no notion of Causation [in Strawson’s sense] to be affirmed (or denied, or even entertained as a possibility)” (1991: 560).

The second way to challenge the realist interpretation of Hume has been offered by Simon Blackburn (1984: 210). To be sure, he does also challenge the received view, by arguing that Hume was far from trying to analyse the *concept* of cause in terms of regular succession. But, on Blackburn's view, the causal realist interpretation is no less mistaken. For, as he claims, the causal realist interpretation enforces on Hume the desire "for a Straightjacket on the possible course of nature: something whose existence at one time guarantees constancies at any later time" (1990: 241). Hume, he notes, warned us exactly against the possibility of apprehending such a Straightjacket: "we have no conception of it, nor any conception of what it would be to have such a conception nor any conception of how we might approach such a conception" (1990: 244). On the positive side, Blackburn takes Hume to be a "projectivist". Our everyday life includes causal behaviour: we draw causal inferences, speak of an effect following with necessity from a cause, and so on. Following Craig (2000: 114), let's call this our "everyday business of 'causalising'". Now, Blackburn argues that Hume was after an *explanation* of this everyday business. But the distinctive interpretative line that Blackburn follows is that Hume was a "projectivist" about causal necessity. When we dignify a relation between events as causal we really "spread" or "project . . . a reaction which we have to something else we are aware about the events"; namely, the regular succession of similar events (1984: 210–11). According to Blackburn, Hume explained the "causalizing" behaviour by this projective theory of necessity. Although the reality that triggers this behaviour "exhibits no such feature [of necessity]", Hume shows how we can still make sense of our "normal sayings" and "our normal operations" with this concept.²³

It is time to bring this long chapter to a close. The "New Hume" debate has certainly advanced our understanding of Hume.²⁴ But it is still inconclusive. What matters is that, even if Hume was not, *very strictly speaking*, a Humean, he made Humeanism possible. In particular, he made RVC possible. In the subsequent chapters, I shall turn my attention to Humeanism and its critics.

Regularities and singular

2 causation

The prospects of singular causation, that is, of causal relations that do not instantiate regularities, will be the main topic of this chapter. We shall focus our attention on Ducasse's critique of Hume and on his defence of singular causation. But, along the way, we shall examine Mill's version of RVC, and Davidson's attempt to reconcile RVC with singular causation.

2.1 Of clocks and hammers

According to Ducasse (1969: 9), Hume offered a definition of the *meaning* of causation: "Causation means nothing but constant conjunction of objects in experience." This definition can be easily demolished. Ducasse (1969: 16) uses the following example to show that constant conjunction is not sufficient for causation. Imagine a man imprisoned since childhood who hears two clocks striking the hours. One of them is faster than the other by one second or so, so that the prisoner hears the first clock striking the hours always a fraction earlier than the other. If Hume's account was correct, then, Ducasse says, the prisoner would be entitled to claim that the "sound of the earlier clock caused the sound of the other". But since this is absurd, Hume's account is not a sufficient condition for causation: correlation does not imply causation. It does not even follow that two correlated events have a common cause. A child's birth is correlated with the ninth reappearance of the moon in the sky since its conception, but there is no common cause for them. Conversely, Ducasse argues, there may be causation

even in the absence of constant conjunction. So constant conjunction is not necessary for causation either. Suppose that a vase gets shattered by being hit with a hammer. This is a case of causation, Ducasse argues, even if there is no constant conjunction between hitting vases with hammers and their shattering. To be sure, Ducasse adds, in order to claim that this sequence of events is causal, one needs also to know that the hammer blow was “the *only* change in the circumstances of the object at that moment” (*ibid.*). This will turn out to be a crucial proviso. Be that as it may, Ducasse thinks that such counter-examples are enough to show that what he took to be Hume’s “definition” of causation is doomed.

Hume’s thoughts about causation are not adequately captured by what Ducasse took to be Hume’s definition. Yet Ducasse’s counter-examples are not in vain. The first set of them – challenging the sufficiency of constant conjunction, or regular succession, for causation – has given rise to an attempt to distinguish, within a Humean setting, between *good* and *bad* regularities, that is, between those regularities that can be deemed causal and those that cannot (such as the constant succession of day by night, or the two-clocks case). We shall devote Chapters 5–7 to this issue. So, for the time being, I shall only examine Mill’s attempt to answer such counter-examples. This will prove relevant to understanding some central differences between Humean and non-Humean accounts of causation. The second set of counter-examples – challenging the necessity of constant conjunction for causation – brought into focus one central issue that separates Humeans from non-Humeans: the existence or not of singular causation, that is, of causation (or causal sequences of events) that does not instantiate regularities.

An important corollary of the discussion of Hume’s views in Chapter 1 is that for Hume (or for Humeans) there cannot be singular causation. In fact, Hume said this much quite explicitly in his first *Enquiry*:

I much doubt whether it be possible for a cause to be known only by its effect . . . or to be of so singular and particular a nature as to have no parallel and no similarity with any other cause or object, that has ever fallen under our observation. It is only when two *species* of objects are found to be constantly conjoined, that we can infer the one from the other; and were

an effect presented, which was entirely singular, and could not be comprehended under any known species, I do not see, that we could form any conjecture or inference at all concerning its cause. . . . both the effect and cause must bear a similarity and resemblance to other effects and causes, which we know, and which we have found, in many instances, to be conjoined with each other. (E: 148)

Kripke (1982: 68) has aptly called this “the impossibility of private causation”. It is the claim that “when the events *a* and *b* are considered by themselves alone, no causal notions are applicable” (1982: 67–8). Causal concepts are applicable only when the relevant events are covered by a regularity.

Ducasse’s own important contribution to the debate was his systematic attempt to offer an account of *singular* causation. Since Ducasse’s own account is built on his criticism of Mill’s views, it will be instructive to examine Mill’s approach first.

2.2 Invariance and unconditionality

Mill defended RVC. But he also improved on it, by offering a more sophisticated version. As he pointed out:

It is seldom, if ever, between a consequent and a single antecedent that this invariable sequence subsists. It is usually between a consequent and the sum of several antecedents; the concurrence of all of them being requisite to produce, that is to be certain of being followed by, the consequent. (1911: 214)

So causal relations relate several factors¹ *C*, *F*, *G*, and so on, with an effect *E* such that the conjunction of all these (call it *CFG*) is sufficient (and, perhaps, necessary) for *E*. Let’s – following Mill – call these factors *positive conditions*. Strictly speaking, Mill adds, negative conditions, namely, the *absence* of some conditions, are also required for the effect *E* to invariably follow. So, for instance, take the case of a fire that followed a short circuit in a house. The fire followed a certain conjunction of positive and negative conditions, which include the short circuit, the presence of oxygen, the presence of inflammable material, the absence of a sprinkler system, the failure of the resident

to notice the beginning of the fire and hence to extinguish it quickly, and so on. If some of the positive factors were *not* part of the conjunction (if, for instance, there was no oxygen present), or if some of the negative factors *were* part of the conjunction (if, for instance, there was a sprinkler system installed), then the fire would not follow the short circuit. In light of this, Mill argues: “The cause then, philosophically speaking, is the sum total of the conditions positive and negative taken together; the whole of the contingencies of every description, which being realised, the consequent invariably follows” (1911: 217). So the real cause is “the whole of these antecedents” (1911: 214). Mill resists the temptation to chop up this totality of factors that precede the effect into two sub-sets: the *cause* of the effect and the mere *conditions* that are needed to be present (or absent) for the cause to produce the effect. No proper part of the full antecedent state *CFG* (e.g. *FG*) is sufficient for *E*, that is, all of the antecedent state *CFG* is necessary for *E*. “. . . we have, philosophically speaking, no right to give the name of cause to one of them exclusively of the others. . . . [For] all the conditions were equally indispensable to the production of the consequent” (*ibid.*). In ordinary life, we do “dignify” a factor, “the one condition which came last into existence”, with the name “cause”. But, Mill adds, this is just “our capricious manner” to “select among the conditions that which we choose to denominate the cause” (1911: 215).

Millian causation, then, is a version of RVC, with the sophisticated addition that in claiming that an effect invariably follows from the cause, the cause should not be taken to be a single factor, but rather the whole conjunction of the conditions that are sufficient and necessary for the effect. Mill (1911: 217–18) does acknowledge that if conditions also include negative factors, then the conditions under which an effect invariably follows cannot be stated properly. His reaction is that the negative conditions can be “summed up under one head, namely, the absence of preventing or counteracting causes”. Then, the negative conditions being the same in all instances (being just the absence of counteracting factors), one can just state the positive conditions as being enough to “make up the whole set of circumstances on which the phenomenon [effect] is dependent” (1911: 218).²

It might be objected that the Millian account is too strong since one can include irrelevant factors in the whole antecedent state that

was sufficient for an effect *E*. So if *CFG* is sufficient for *E*, then so is *ACFG* where *A* might be totally irrelevant, for example, that someone was crossing the road when the fire started. But this can be easily dismissed on the grounds that this factor was not necessary for the fire. That is, in general *CFG*-and-*not-A* is sufficient for the effect *E*. Another objection might be that Mill's denial of the difference between causes and conditions might lead him to accept *trivially* relevant causal factors. So suppose that a person died after drinking arsenic. Why shouldn't we include in the conditions of her death the fact that she was human and not, say, a plant, or the fact that she was a woman and not a man, or indeed the fact that she was alive before her death? Millians are on safe ground here, if they accept Mackie's (1974: 63) notion of a "causal field". This is the *context* in which the conditions of an effect occur. The causal field should not be taken to be part of the conditions that are sufficient for the effect in the sense that it is the *background* "against which the causing goes on" (*ibid.*). This background would be there even if the specific conditions that are sufficient for the occurrence of the effect were absent.

Now, Mill improved on RVC in one more respect. He did strengthen it against the first type of counter-examples that were suggested by Ducasse (see section 2.1). For Mill, regular association (although necessary for causation) is not, on its own, sufficient for causation. The night has always followed the day, but it's not caused by it. He (1911: 221) noted: "When we define the cause of anything . . . to be 'the antecedent which it invariably follows', we do not use this phrase as exactly synonymous with 'the antecedent which it invariably has followed in our past experience'." Rather, we mean to assert that "as long as the present constitution of things endures, [the antecedent] always *will* be [followed by the consequent]" (*ibid.*).

Mill goes on, albeit in a footnote, to tie this last claim to the existence of laws of nature. Those invariable successions are causal which constitute laws of nature. For only those cannot be "terminated or altered by natural causes" (1911: 221). The sequence of day and night might be so terminated, that is, the possibility of its termination is consistent with "the ultimate laws of nature", if, for instance, an opaque body comes in between the sun and the earth. As we shall have the opportunity to see in Chapter 5, this link between causation and laws of nature has become a prominent feature of the

Humean approach to causation. But Mill added more by way of explication. An invariable succession of events is causal only if it is “unconditional”, that is, only if its occurrence is *not* contingent on the presence of further factors which are such that, given their presence, the effect would occur even if its putative cause was not present. What Mill has in mind is that an invariable succession is unconditional if it is robust enough not to be explained away by the presence of further factors (cf. 1911: 222). A clear case in which unconditionality fails is when the events that are invariably conjoined are, in fact, effects of a common cause.³ In his own words:

There are sequences, as uniform in past experience as any others whatever, which yet we do not regard as cases of causation, but as conjunctions in some sort accidental. Such, to an accurate thinker, is that of day and night. The one might have existed for any length of time, and the other not have followed the sooner for its existence; it follows only if certain other antecedents exist; and where those antecedents existed, it would follow in any case. (*Ibid.*)

Here is an example. Drops in the barometric readings are correlated with subsequent storms. But this sequence is not causal, because it is not unconditional. If we take into account the further factor of the *fall of the atmospheric pressure*, it is clear that the storm would follow, even if the barometric reading had not dropped. Once more, Mill’s seminal suggestion was destined to become a standard empiricist way to distinguish between invariable successions that are causal and those that are not. As we shall see in section 9.3, this idea of unconditionality is currently expressed by the claim that a correlation is causal if it cannot be *screened off*. Yet the effectiveness of Mill’s proposal depends on the ability to assess certain counterfactual claims, such as “if x hadn’t happened, then y would still have followed”. This suggests that there is more to causation than actual regular succession. In any case, Mill (1911: 222) was clear: “Invariable sequence, therefore, is not synonymous with causation, unless the sequence, besides being invariable, is unconditional.” And: “We may define, therefore, the cause of the phenomenon to be the antecedent, or the concurrence of antecedents, on which it is invariably and *unconditionally* consequent” (*ibid.*).

2.3 Agreement and difference

It is important to examine how Mill proposed to test whether a succession of events is causal. Apart from being interesting in its own right, this issue will be relevant to what will follow in the next section concerning *singular causation*.

Mill's methods are known as the Method of Agreement and the Method of Difference (cf. 1911: 253–6). Suppose that we know of a causal factor *C* and we want to find out its effect. We vary the factors we conjoin with *C* and examine what the effects are in each case. Suppose that, in a certain experiment, we conjoin *C* with *A* and *B* and what follows is *abe*. Then, in a new experiment, we conjoin *C* not with *A* and *B* but with *D* and *F* and what follows is *dfe*. Both experiments agree only on the factor *C* and on the effect *e*. Hence the factor *C* is the cause of the effect *e*. *AB* is not the cause of *e* since the effect was present even when *AB* was absent. Nor is *DF* the cause of *e*, since *e* was present when *DF* was absent. We can apply a similar method if we want to know what the cause of a given effect *e* is. Suppose we observe the effect *e* in different combinations with other effects, such as *abe* and *dfe*. If we can produce experiments such that in one of them the antecedent of *abe* is *ABC*, while in the other the antecedent of *dfe* is *DFC*, we can conclude, by reasoning similar to the above, that the cause of *e* was *C*. This is then the Method of Agreement. The cause is the common factor in a number of otherwise different cases in which the effect occurs. As Mill (1911: 255) put it: "If two or more instances of the phenomenon under investigation have only one circumstance in common, the circumstance in which alone all the instances agree is the cause (or effect) of the given phenomenon."

The Method of Difference proceeds in an analogous fashion. Where in the Method of Agreement we knew the cause (or the effect) we wanted to investigate, in the Method of Difference we don't have such knowledge. So suppose that we run an experiment and we find that an antecedent *ABC* has the effect *abe*. Suppose also that we run the experiment once more, this time with *AB* only as the antecedent factors. So factor *C* is absent. If, this time, we only find the part *ab* of the effect, if that is, *e* is absent, then we conclude that *C* was the cause of *e*. Conversely, if in an experiment we get an effect *abe* from an antecedent state *ABC*, and we want to find the cause of *e*, we should try to produce another experiment with

effects *ab*. If we find out that, in this new experiment, the antecedent factors are *AB*, that is if, when the effect *e* is absent, so is the factor *C*, then we can conclude that *C* was the cause of *e*. On the Method of Difference, then, the cause is the factor that is different in two cases, which are similar except that in the one the effect occurs, while in the other it doesn't. In Mill's words:

If an instance in which the phenomenon under investigation occurs, and an instance in which it does not occur, have every circumstance in common save one, that one occurring only in the former; the circumstance in which alone the two instances differ is the effect, or the cause, or an indispensable part of the cause, of the phenomenon. (1911: 256)

Of the two methods, the Method of Agreement can show only that a sequence of events is invariable: that whenever *C* happens, *e* follows. It cannot show that this sequence is unconditional. This is because it is hardly ever possible to ascertain that two states agree on all but one factors. It may well be that two states *ABC* and *DFC*, whose effects are *abe* and *dfe*, respectively, agree not just on *C* but also on other unknown factors *K*, *L*, *M*, and so on. And these unknown factors may well be just those that bring about the effect *e*. The Method of Difference, on the other hand, stands a much better chance to establish unconditionality, and hence causation, for all the Method of Difference requires is the careful removal of a single factor *C* from a state *ABC*. If the effect *e* is thereby removed too, it can be safely concluded that *C* caused *e*, since if it didn't, and if some of the other remaining factors did, the effect *e* would still be present. Now, as Mill (1911: 256) notes, a situation such as the above can take place in a controlled experiment. Suppose that a state *AB* exists before the controlled experiment takes place and that the effect is *ab*. Suppose, further, that the experiment introduces a further factor *C*, and a new effect *e* is observed. Clearly, *C* caused *e* because *C* was the only difference between *AB* (where effect *e* was absent) and *ABC* (where the effect *e* was present). For instance, let's say that we want to find out whether a change in the length of the rod of a pendulum causes its period to change. We alter the length, while keeping all else as it was, and see whether the period changes – it does.

So it is only in the environment of a controlled experiment that we can draw relatively safe conclusions about causal connections between factors. When we make natural observations of two states *ABC* and *AB*, it is hardly ever possible to be sure that the only difference between them is *C*. Hence it is hardly ever possible to draw safe causal conclusions from natural observations by applying the Method of Difference. But even in a controlled experiment, where we carefully prepare the situation in such a way that we control for all factors but one, it is still possible, as Mill (1911: 257) notes, that the effect was produced not by the newly added factor *C* but “by the means employed to produce the change” from the previous state *AB* to the new *ABC*. So, unless further careful experiments preclude this possibility, it is not entirely safe to draw causal conclusions even by means of the Method of Difference.

In any case, Mill is adamant that his methods work only if certain metaphysical assumptions are already in place. First, it must be the case that events have causes. Secondly, it must be the case that events have a limited number of possible causes. In order for the eliminative methods he suggested to work, it must be the case that the number of causal hypotheses considered is relatively small. Thirdly, it must be the case that same causes have same effects, and conversely. Fourthly, it must be the case that the presence or absence of causes makes a difference to the presence or absence of their effects. Indeed, Mill (1911: 255) made explicit reference to two “axioms” on which his two Methods depend. The axiom for the Method of Agreement is this:

Whatever circumstances can be excluded, without prejudice to the phenomenon, or can be absent without its presence, is not connected with it in the way of causation. The casual circumstance being thus eliminated, if only one remains, that one is the cause we are in search of: if more than one, they either are, or contain among them, the cause . . . (*Ibid.*)

The axiom for the Method of Difference is: “Whatever antecedent cannot be excluded without preventing the phenomenon, is the cause or a condition of that phenomenon: Whatever consequent can be excluded, with no other difference in the antecedent than the absence of the particular one, is the effect of that one” (1911: 256).

What is important to stress is that although only a pair of (or even just a single) carefully controlled experiments might get us to the causes of certain effects, what, for Mill, makes this inference possible is that causal connections are embodied in invariable successions. It is because causation is so embodied, that the Methods of Agreement and Difference justify us in drawing causal conclusions. He couldn't be clearer on this: "The Method of Agreement stands on the ground that whatever can be eliminated is not connected with the phenomenon by any law. The Method of Difference has for its foundation, that whatever cannot be eliminated is connected with the phenomenon by a law" (1911: 256).

2.4 Single difference

Ducasse's point of departure is that Mill's Methods, and especially the Method of Difference, freed of the requirement of invariability, can provide a *definition* of causation in a single instance (or in a single experiment). According to him, Mill wrongly regards "single difference as a 'method' for the roundabout ascertainment of something other than itself, viz. of invariable sequence; instead of, and properly, regarding it as the very definition of cause" (1968: 7). So, Ducasse suggests, "the requirement of invariability be cast to the winds, for we no longer need it" (1969: 21).⁴ For, "causation is directly concerned with single cases, not with constant conjunctions" (*ibid.*).

Before we examine Ducasse's definition in some detail, a couple of general notes are in order. First, his aim is to offer an analysis of the concept of causation, that is, a definition of it, which does justice to the proper and typical uses of this concept in everyday discourse (cf. 1968: 1–2). In particular, his attempt is to define causation in non-causal terms. So Ducasse's project can be seen as a version of a singularist *reductive* account of causation. Secondly, one thing that Ducasse sets straight right away is that the relata of a causal relation are *events*. Hume spoke of "objects" as being causes and effects, but objects (e.g. a tree, or a table or a person), or substances, cannot really be the relata of causal relations. It was not the tree, but the *falling* of the tree that caused the injury to the lumberjack under it. It was not the hammer, but the *hitting* of the vase with the hammer that caused its smashing. But what are events? This is a vexed philosophical

question to which I shall briefly try to sketch a couple of answers in section 2.6. For the time being, let us only note what Ducasse's view is. He (1969: 52–3) takes an event to be a *change* or a *state* of an object. But what changes (or remains the same) is a *property* of an object. It is the change of the position of the sun in relation to the building that caused the change of the shape of its shadow. So it seems fair to say that Ducasse takes events to involve changes in the properties of things. When we talk of a particular event (an *event-token*), apart from specifying the particular object in which the change happens, we individuate it by adding the specific place and the specific time in which this change takes place. But we can also talk of *event-types*, that is, of changes in objects generically understood. An event-type might be a smashing of a vase, and an event-token (an individual event of this type) might be the particular smashing of a particular vase that happened in a particular spatiotemporal location. With this in mind, let's go straight to Ducasse's definition (1968: 3–4, cf. also 1969: 54–7), which I shall state in full:

(S)

Considering two changes *C* and *K* (which may be either of the same or of different objects), the change *C* is said to have been sufficient to, i.e., to have caused, the change *K*, if:

1. The change *C* occurred during a time and through a space terminating at the instant *I* at the surface *S*.
2. The change *K* occurred during a time and through a space beginning at the instant *I* at surface *S*.
3. No change other than *C* occurred during the time and through that space *C*, and no change other than *K* during the time and through the space of *K*.

More roughly, . . . we may say that the cause of particular change *K* was such particular change *C* as alone occurred in the immediate environment of *K* immediately before.

What is immediately clear is that Ducasse offers a definition of *singular causation*. There is no supposition of regular association between *C* and *K*. All that is required for *C* to be the cause of *K* is a *single occurrence* of *C* (and of course the non-occurrence of other changes after *C* and before *K*). The relata *C* and *K* are

“concrete individual events” (1968: 6), that is, event-tokens. What is, then, the hallmark of singular causation? It is the claim that what makes a sequence of event-tokens *c* and *e* causal is an *intrinsic* feature of this sequence, which in no way depends on things that happen at other places and in other times. To see how different this view is from RVC, consider this. Defending RVC, Carnap (1974: 201–2) noted that the causal relation behaves very differently from other relations. In most cases, if we want to determine whether a certain relation (e.g. *x* is taller than *y*, or *x* occurs immediately after *y*) holds between two objects (or events) *a* and *b*, we examine directly these two objects (events). We do not have to examine any other objects (events) of the same type. But, on RVC, when it comes to ascribing a *causal* relation between event-tokens *c* and *e*, there is nothing in the *particular* pair of events that determines whether they are causally related to each other. According to RVC, whether a particular sequence of events will be deemed causal depends on whether it instantiates a regularity. That is, it depends on whether event-tokens *c* and *e* fall under event-types *C* and *E* such that all events of type *C* are regularly associated with (or, regularly followed by) events of type *E*. This means that whether a sequence is causal depends on things that happen in another time and place in the universe. Unlike RVC, Ducasse takes causation to be fully definable without any appeal to “the notion of law or uniformity” (1969: 41) and with reference only to what happens in an individual case.

It is noteworthy that Ducasse does allow that there are causal regularities in nature, expressed by general statements linking event-types. However, he takes them to be “corollaries” of causation and not a definition of it. As he (1969: 21), in a rather uncharitable moment, said: “To have mistaken it [the constant conjunction] for the latter [the definition of causation] was Hume’s epoch making blunder, which has infected directly or indirectly every discussion of causation since.” For Ducasse, in offering his own regularity theory of causation, Mill just “borrow[ed] the blunder” (1968: 7). So statements of regularities or laws are, for Ducasse, *generalizations* of singular causal facts. When individual events that occur in causal relations exhibit likeness to others, they are grouped together into event-types and a general causal statement. The general statement is causal not because it states a constant conjunc-

tion between event-types, but because it is grounded in a class of *causal* facts about resembling event-tokens (cf. 1968: 6–7).⁵

Is Ducasse's definition (S) adequate? Ducasse himself (1969: 77) describes the major objection to it in an excellent way. One might object that if we take (S) seriously, then "the distinction is altogether lost between consequence and mere sequence". Humeans are able to distinguish cases in which two event-tokens merely follow each other from cases in which they are causally connected by imposing the extra requirement that the causal sequence instantiate a regularity. But singularists, like Ducasse, seem unable to offer a good mark of causality. (S) offers only a sufficient condition for causation. There are reasons to think that he would also need (S) to offer a necessary condition for causation. But let's leave that to one side. For it can be questioned whether (S) offers even the intended sufficient condition for causation. Suppose that a small fraction of time before I let go of the stone that I am leisurely holding in my hand, I also, quite unconsciously, flick my hand upwards. The two events are spatiotemporally contiguous in the way demanded by (S). Besides, no change other than the flicking of my hand upwards occurred before my letting go of the stone. So all conditions of (S) are satisfied. Yet we cannot say that the cause of my letting go of the stone was the flicking of my hand upwards. One event followed the other but it was not caused by it. So there is a sense in which Ducasse's account makes accidents (or coincidences) look like genuine causings.

A related (but different) objection is this: (S) seems to be *vacuous*. It requires that for C to be the cause of K, C must be the *only* change in K's environment before its occurrence, and yet it seems to make no sense to say that there was just a single change in the immediate environment of K before K occurred. The point is not epistemic – although the epistemic dimension is also important. That is, the point is not how we could possibly identify the only change C in K's environment, even if there was such a single change. The issue is metaphysical, for, strictly speaking, there is no such single change C in K's environment. Take the standard example of the smashing of the vase. What was the single change in the environment of the vase before its smashing? There is no such thing. The position of the hammer changed, the physical properties (e.g. the kinetic energy) of the hammer changed, the surrounding

air-molecules changed their positions and properties, the shadow that the hammer cast changed (let us suppose that the hammer was illuminated) and so on. But if there is no single difference in the vase's environment, then there is no cause of the vase's smashing, according to (S). This is absurd, and hence (S) must be wrong. Take another case. Suppose that simultaneously with the hammer blow, the person with the hammer in his right hand snapped the fingers of his left hand. How are we to exclude this change in the vase's environment from being the cause (or part of the cause) of the smashing of the vase? A natural thought would be to say that the snapping of the fingers was *not* causally relevant to the breaking. But then we would need an account of what changes in the effect's environment were *causally relevant* to its occurrence, and this would render (S) circular, defying Ducasse's intention to define causation in non-causal terms. For, then, the cause *C* of *K* would be the single causally-relevant-to-*K* difference in *K*'s environment.

This last point seems quite compelling and can be further strengthened. Ducasse (1969: 69–70) does consider a relevant objection: in identifying the cause of the smashing of the vase, why shouldn't we also take into account changes in the vase's environment (broadly understood) such as the movement of the moon before the breaking, or the coming-and-going, of the waves in the ocean and so on? He replies that such changes are irrelevant because "they occurred during the existence of the vase without its breaking" (1969: 70). That these changes did indeed occur during the vase's existence is surely the case. Still, it does not follow that they did not causally contribute to its breaking. In order to show this, Ducasse would first need to show that these actual changes were *causally irrelevant* to the breaking of the vase. Indeed, he does state the "Postulate of Non-Interference of Nature" (1969: 71), which is necessary for singular causal claims to be grounded: "The circumstances which were not observed remained during the investigation causally equivalent in respect of the effect investigated."

This might well be an admission of defeat. For one, it is questionable whether this postulate can be defended. It sounds overly *ad hoc* to assume that only the observed changes were causally relevant to the effect. For another, even if it was correct, what this postulate makes clear is that (S) can be an adequate account of causation only if there is a prior distinction between causally relevant and causally

inert changes. In order, however, to draw this distinction, singularists need to stop looking at sequences of individual events and start looking at regularities, that is, at relations between event-types. This is, in effect, what Mill did (see section 2.3). That is, to say the least, singularists have to behave *as if* they were Humean.

But it may be that Ducasse could avoid commitment to regularities, by importing a counterfactual element in his account of causation. So Ducasse might just be able to explain what it is for an event x to be causally irrelevant to an event y , by relying on counterfactual conditionals of the form: “if x hadn’t happened, then y would still have happened”. This would be an important improvement, which would, once more, emphasize the need to import counterfactuals in the analysis of causation. As we have already seen, this need also arises for the Millian regularity account. Since we shall deal with the role of counterfactuals in causation in section 3.3, I shall only raise the worry (to be assessed later on) that counterfactuals may not be able to be accounted for without some resort to regularities.⁶

2.5 Observing causation

There are a couple of aspects of Ducasse’s account that need to be further examined, since they have an independent interest for the singularist approach to causation. For Ducasse, the causal relation is such that there is a necessary connection between cause and effect. Besides, he takes this necessary connection to be directly *observable* (cf. 1969: 58–9). Both of these claims are very striking. They fly in the face of the well-entrenched belief that Hume proved beyond doubt that there is no observation of necessary connections in nature. To be sure, Ducasse does not think that there is a *sensation* of necessary connections, like a sensation of sound or of colour. Still, his thought is that causation, being a relation among observable entities, is no less observable than other such relations. He (1968: 9) says: “. . . its presence among . . . events is to be observed every day. We observe it whenever we perceive that a certain change is the only one to have taken place immediately before, in the immediate environment of another.”

But are things so easy? We already saw in the previous section that Ducasse’s definition (S) is deeply problematic. Besides,

Ducasse has not told us why a sequence of two individual events that satisfy his definition (S) involves a necessary connection between these two event-tokens. That it does so cannot just be a matter of definition, for this would beg the question against the Humeans. Even if we took the cause of an individual event e to be the *total* antecedent state of the universe U , the sequence $\{U, e\}$ could be an accident of a cosmic scale. Nor, of course, is U observable. But suppose we were to grant, for the sake of the argument, that a necessary connection *is* a connection between event-tokens that satisfy Ducasse's definition (S). Does it follow that the so-called necessary connection is *observable*? Not really. The fact that two event-tokens c and e (e.g. the hitting of the porcelain vase with a hammer and its smashing) are observable does not imply that *relations* between them are observable. Take a dog and an amoeba. They are related by the following relation: a dog has more cells than an amoeba. But the relation *has more cells than* is *not* observable, although the relata are.

Ducasse seems to think that the relation of necessary connection *is* indeed observable because we can (and do) directly perceive *causings*. Suppose, for instance, that we directly perceive a branch of a tree being broken by a person who exerted some force on it. Ducasse (1969: 151) argues that we thereby directly perceive that the branch was *caused* to break. Surely, this is too quick. We do perceive the branch, its bending, its actual breaking, and so on. But from this it does not follow that we also directly perceive something else; namely, the causing. In any case, we don't perceive the breaking *as* causing. Compare: from our perceiving of a glass of water it does not follow that we perceive this glass of water *as* a glass of H_2O . We may (rightly in the case of water) *infer* that what we have perceived is a glass of H_2O . Similarly in the case of causings. Even if the act of breaking *is* an act of causing, we can move from the one to the other only by means of an inference.⁷

There is, however, an interesting (and popular) argument in favour of the observability of singular causings that needs to be attended to. Although it is anticipated by Ducasse (1969: 151), it is fully developed by G. E. M. Anscombe (1971) and has a prominent place in Stroud's (1977: 230–1) defence of the view that the concept of causation is intimately linked with the concept of necessary connection. I shall call it *the argument from causal verbs/concepts*. Ducasse points out that

“common verbs which are verbs of causation” embody causal relations. Examples of such verbs are “to bend, to corrode, to push, to cut, to scratch, to break, to kill, to transform, to remind, to motivate, to irritate, to ignite, to create, to incite, to convey etc.” (1969: 150). So we are told that when one asserts that, for instance, a particular tree branch *bent* after having had pressure exerted on it, by the very use of the verb *to bend*, one makes a causal claim and, if this claim is indeed correct, one has thereby directly perceived the tree *being caused to bend*. To bend is to cause to bend, to scratch is to cause to scratch, and so on. Anscombe ([1971] 1993: 93) reinforced this argument by asking us to imagine how the very general concept of cause enters a language. It does, she suggests, by being an abstraction of many other ordinary *causal* concepts. She says:

. . . the word “cause” can be *added* to a language in which are already represented many causal concepts. A small selection: *scrape, push, wet, carry, eat, burn, knock over, keep off, squash, make* (e.g., noises, paper boats), *hurt*. But if we care to imagine languages in which no special causal concepts are represented, then no description of the use of a word in such languages will be able to present it as meaning *cause*. ([1971] 1993: 93)

Anscombe’s point is this: since our language is infested with causal concepts such as the above, and since we can apply them to describe correctly several events that fall under them, there is no further mystery in the claim that we directly perceive causings. For the things to which these concepts apply (e.g. the cat’s making a noise) are *causings*. When we learn to report such things from having observed them, we have thereby learned to report *causings* from having observed them.⁸

This argument has been developed further by Cartwright (1993), who outlines a certain account of the semantics of the concept of cause. She (1993: 426) suggests that the relationship between the concept of cause and specific causal concepts such as those suggested by Anscombe can be usefully seen as the relationship between *the abstract and the concrete*: “. . . To call a lapping of the milk or a de-exciting of an atom a causing is to give a more abstract description of it.” So when we say that, for instance, the blow with the hammer *caused* the breaking of the vase, and when we just talk of the *breaking*

of the vase, we refer to the very same act with a more abstract concept (caused) and a more concrete one (breaking), the concrete concept being an instance of the abstract. The abstract concept, Cartwright notes, has excess content over its particular concretizations; hence it cannot be captured by “a big disjunction across its manifestations” (*ibid.*). Nor, she adds, “need it supervene on them” (*ibid.*). Yet it’s not as if we introduce a wholly new concept when we describe an event as an instance of *causing*. Rather, we move towards a more abstract level of description.⁹ Consequently, she says, when we, for instance, observe the breaking of a vase, we also observe the causing: we don’t infer it; we directly observe it.

How compelling is, then, the argument from causal verbs (concepts)? A lot will turn on how exactly we understand the semantics of abstract concepts. One thing that seems relevant is this. Even if we were to grant that the concept of cause is an abstract concept whose manifestations are instances of breaking, pushing, pulling, creating, and so on, it seems that this abstract concept has implications that (a) exceed those of its manifestations; and (more importantly) (b) impair the claim that when we observe an event such as the above we *ipso facto* observe a causing. Talk in terms of causes seems to imply certain counterfactual conditionals, which are not implied by talk in terms of its so-called manifestations. That the blow with the hammer broke the vase does not necessarily imply that if the vase had not been struck with the hammer, it would not have been broken. However, to say that it was the blow with the hammer that *caused* the vase to break seems to imply that if the vase had not been struck with the hammer, it would not have been broken. If this is so, then two things follow. First, the excess content that the abstract concept of cause has might well need an appeal to regularities to be specified. For, it may well be the case that assessing the truth of the relevant counterfactual conditionals might require taking account of the relevant regularities. Secondly, if causation involves counterfactuals, then it is clear that causation cannot be observed in single sequences of events, even when these sequences of events are directly observable. For, whatever else they are, counterfactuals are not observable (cf. Mackie 1974: 142). Hence, there seem to be at least *prima facie* reasons to think that the argument from causal verbs/concepts is not compelling.

2.6 Phantom laws

In trying to analyse the *logical form* of singular causal statements such as

the short circuit caused the fire

Davidson (1967) suggested that (a) they should be seen as linking individual (particular) events and (b) they should be analysed as follows:

There exist unique events c such that c is a short circuit and e such that e is a fire, and c caused e .

Taking F to stand for *short circuit* and G for *fire*, the singular causal statement has the form

The unique event c , such that Fc , caused the unique event e , such that Ge (cf. Davidson [1967] 1993: 80).

Statements such as the above are, for Davidson, genuine causal statements. They can truly state that one event caused another. There is, however, the further question of whether such statements, which on the face of them commit us to singular causation, are complete. Wouldn't we, for completeness, require that such statements be deduced from (or instantiate) relevant lawful generalizations? Making a sharp distinction between an event and its *description*, Davidson argues:

What is partial in the sentence "The cause of this match's lighting was that it was struck" is the *description* of the cause; as we add to the description of the cause we may approach the point where we can deduce, from this description and laws, that an effect of the kind would follow. ([1967] 1993: 81)

It is difficult to analyse carefully what Davidson says here, without going into his theory of events. So I'll offer a few standard thoughts. Davidson takes events to be spatiotemporal *particulars*, which can be described in different ways. Specifically, they can be referred to by means of singular terms and definite descriptions. A

certain event, for instance, can be properly referred to as the stealing of Jones's car. But it can also be referred to as the stealing of the red Ferrari with number plates "LUV 0001". Events, Davidson stresses, should not be confused with their descriptions. Their descriptions can be partial, perspectival or incomplete. The very same event can be referred to by means of different descriptions. For instance, the event that is described as *the earthquake that shook Athens up at noon on 7 September 1999* can also be referred to as *the event that was in the headlines of all of Athens's newspapers on 8 September 1999*. Or, even as *the event that led to the greatest number of deaths on a single occasion during peacetime in Athens*. What Davidson points out is that it is the events themselves that enter into causal relations, and *not* their descriptions. Descriptions refer to them and report them, but don't have a causal role. Take the causal statement: *the 1999 Athens earthquake caused a great wave of panic among Athenians*. No matter which description of the antecedent event *c* we choose, it is still the case that *c* caused *a great wave of panic among Athenians*. In particular, it is the case that *the event that was in the headlines of all of Athens's newspapers on 8 September 1999 caused a great wave of panic among Athenians*. It follows from Davidson's reading that causal statements of the form *c caused e* are *extensional*: their truth-value does not change, if we substitute for the descriptions of the two events *c* and *e* new descriptions of them.

So an intuitive way to think of Davidson's approach to events is to think of them as particular happenings in particular spatio-temporal locations that can be referred to by different descriptions. It's not surprising then that Davidson ([1967] 1993: 82) notes that it is "a confusion" to think that "every deletion from the description of an event represents something deleted from the event described". In light of this, we can see how Davidson takes it to be the case that singular causal statements, if true, are also complete. For the events themselves that enter into causal relations are complete. Only their descriptions can be incomplete. But then we can also see how he claims to be able to reconcile the Humean RVC and its singularist opposition. For when we pick the descriptions of the events that enter a causal statement, the descriptions may be such that they entitle us to *deduce* the singular causal statement from a lawlike statement together with the assumption that the

events referred to in the statement occurred. So we can subsume the singular causal statement under a causal law. His suggestion (1967: 83) is that if “*c* causes *e*” is true, then there *must* be a law from which this causal statement follows, even if this law is unknown to those who use the singular causal statement, and even if the law is not stated in the vocabulary of the singular causal statement. The existence of this law captures the sense in which Humeans say that causes are *necessary and sufficient* for their effects.

If Davidson is right, then, as he notes, the Humeans and the singularists are not in conflict: “the reconciliation depends, of course, on the distinction between knowing there is a law ‘covering’ two events and knowing what the law is: in my view, Ducasse is right that singular causal statements entail no law; Hume is right that they entail there is a law” ([1967] 1993: 85). As with many suggested compromises, the parties for which the settlement is proposed are likely to protest that it does not do justice to them. In Davidson’s case at least, they have. To be sure, the Humeans seem to be better off than the singularists, if they accept the reconciliation. For Davidson’s concession to the Humeans is metaphysical, while his concession to the singularists is epistemic. If his argument is right, then the *truth conditions* of singular causal statements are fully specified by reference to laws. In fact, Davidson suggests that even if we are unable to formulate a law of the suitable form, there is no reason for despair. For when we use true statements of the form “*a* caused *b*”, we are committed that “there are descriptions of *a* and *b* such that the result of substituting them for ‘*a*’ and ‘*b*’ is entailed by true premises of the form [the relevant law] and [initial conditions]; and the converse holds if we put suitable restrictions to the descriptions” ([1967] 1993: 84).

What then makes singular causal statements true is the existence of some regularities or laws. All causation is nomological: *c* causes *e* iff there is a *law* that connects events like *c* with events like *e*. From a metaphysical point of view, then, causation is what Humeans take it to be. What Davidson ([1967] 1993: 85) concedes to the singularist is an epistemic point, but even this concession is limited: “. . . it does not follow that we must be able to dredge up a law if we know a singular causal statement to be true; all that follows is that we know there must be a covering law.” Humeans might not have to deny this. But, as we shall see in some detail in section 8.3.2, a natural Humean

objection to it – attributable to Hempel – is that the mere claim that there is a law but we don't need to know it, is tantamount to the claim that there is a treasure hidden somewhere here, but we don't need to find it. To use Scriven's (1975: 8) memorable expression, all that Davidson's compromise assures us of is the existence of "phantom laws". For the time being, let me stress that the epistemic concession to the singularists is unlikely to satisfy them either. For they want to have an account of singular causation as it is in the world. Davidson's compromise does not offer this. It offers a Humean account of causation as it is in the world. That we can have genuine causal knowledge by means of true singular causal statements is not much of a comfort to the singularists. They would be more willing to capitalize on the fact that we know singular statements to be true in order to show that, the truth conditions of causal statements being accessible, they should *not* be given by any reference to laws at all; nor even by reference to "phantom laws". This point is in fact pressed by Woodward (1986: 270–71) who, in defending singular causation against Davidson's argument, points out that Davidson makes a mystery of the fact that although true singular causal statements do explain why an event happened, the laws that, on his view, are part of what does the explaining, are (or may be) inaccessible.

The thought that, for singularists, there must be singular causation as it is in the world is reinforced by Anscombe's famous comment on Davidson's compromise:

Meanwhile in non-experimental philosophy it is clear enough what are the dogmatic slumbers of the day. It is over and over again assumed that any singular proposition implies a universal statement running "Always, when this, then that"; often assumed that true singular causal statements are derived from such "inductively believed" universalities. Examples indeed are recalcitrant, but that does not seem to disturb. Even a philosopher acute enough to be conscious of this, such as Davidson, will say, without offering any reason at all for saying it, that a singular causal statement implies *that there is* such a true universal proposition – though perhaps we can never have knowledge of it. Such a thesis needs some reason for believing it! "Regularities in nature": this is not a reason.

([1971] 1993: 104)

There are also other reasons to challenge Davidson's views. Most of them turn on his account of *events*. Some philosophers deny that the causal relata are events. Mellor (1995), for instance, argues that causal statements relate *facts*, where facts may be seen as whatever true propositions express. Hence, Mellor rejects Davidson's analysis of the logical form of singular causal statements, and forfeits the need to offer an analysis of events.¹⁰ Others, however, accept that the causal relata are events, but offer a different analysis of events. A dominant alternative view, attributed mainly to Kim (1971, 1993), is that events are *exemplifications of properties* by objects at particular times. So an event is a triple $[x, P, t]$, which states that the property P is exemplified by the object x at time t . An advantage of this account over Davidson's is that it makes clear how *properties* can be causally efficacious. For instance, when we say that it was the excessive weight of person S that caused the plastic chair to break, we mean that it was the *excessiveness* of the weight that caused the breaking. In fact, this view of events has been adopted by Armstrong (1983: 94–5, Heathcote & Armstrong 1991: 67–8), in an attempt to show how singular causation can be reunited with causal laws in a better way than the one offered by Davidson. Since this way depends on Armstrong's view of laws as relations among universals, we shall offer a discussion of it in section 6.3.3. But all this is only a tiny fragment of the different views of what events are (cf. Mackie 1974: Ch. 10, Bennett 1987, Hausman 1998: 19–21, and the references therein). The notion of an event seems to resist an uncontroversial philosophical analysis. So one might have to take very seriously Hausman's suggestion: "Even if there were a promising theory of events, one should pause before erecting a theory of causation on the shaky foundations of a theory of events" (1998: 21).

To sum up, despite its intuitive appeal and the ingenious efforts of its advocates, the possibility of genuinely singular causation, that is, of causation which is not an instance of a regularity, has not yet been established.¹¹ Davidson's attempted reconciliation would be more acceptable to Humeans, yet both Humeans *and* singularists contest its soundness. But it is now time to cash in a promissory note that was issued in several places in this chapter: *the role of counterfactuals in causation*.

3 Causation and counterfactuals

In the first *Enquiry*, after the statement of the first definition of causation, which, as the reader might recall, was a Regularity definition, Hume (E: 146) added the following, *prima facie* puzzling, remark: “Or in other words, where, if the first object had not been, the second never had existed.” But these are not merely “other words”. They offer a distinct definition of causation. They define causation not in terms of actual regularities, but in terms of a counterfactual dependence of the effect on the cause: the cause is rendered counterfactually necessary for the effect. Almost everybody agrees that counterfactual conditionals play an essential role in causation. But there is important disagreement on what exactly this role is. In this chapter, I shall examine in some detail two representative views of the role of counterfactuals in causation, one associated with Mackie and the other with Lewis. Finally, in section 3.4, I shall introduce the reader to two notable recent attempts to improve on earlier counterfactual theories.

3.1 The meaning of causal statements

Mackie makes a nice distinction between two sorts of question one can ask about RVC. First, does it capture the *meaning* of singular causal statements? Secondly, does it capture what constitutes causation as it is in the objects? (cf. 1974: 77). Although he himself offers an improved version of RVC (which we shall discuss in section 3.2), Mackie answers both of the foregoing questions in the negative. Let

us, in this section, see his reasons for the negative answer to the first question.

What is the meaning of singular causal statements? When, for instance, we say that the fire in the house was caused by the short circuit, what exactly do we assert? Mackie argues that a causal statement of the form “*c* caused *e*” should be understood as follows:

c was necessary in the circumstances for *e*,

where *c* and *e* are distinct event-tokens. Necessity-in-the-circumstances, he adds, should be understood as the following counterfactual assertion:

if *c* hadn’t happened, then *e* wouldn’t have happened (cf. 1974: 31).

Mackie repeatedly stresses that counterfactuals such as the above are involved in his analysis of the concept of singular causation. His main objection to Humeanism is that it fails to capture the *meaning* of singular causal statements. Like Ducasse (see section 2.3), he takes singular causal statements to be prior to general causal statements. Mackie views the latter as “quantified variants of the corresponding singular ones” (1974: 80). So, for instance, the general statement “Heat causes gases to expand” is taken to be a generalization of the relevant singular causal statements and is such that the meaning of *causes* in the general statement is the same as the meaning of *caused* in the singular causal statement “The heating of this gas caused it to expand”. This meaning, Mackie says, is fixed by the appropriate counterfactual conditionals (cf. also Mackie 1974: 267–8, 270). So a lot turns on what he thinks about counterfactuals. In particular, an important issue is under what circumstances a counterfactual conditional can be asserted.

3.1.1 A journey to possible worlds

Take the statement “If this match had been struck, it would have lit”. What, if anything, makes it true? Mackie has developed his views on counterfactuals in several places (cf. 1973: Ch. 3, 1974).

His main thought is that counterfactual statements such as the above are not, strictly speaking, true or false: they do not describe, or fail to describe, “a fully objective reality” (1974: xi). Instead, they are reasonable or unreasonable assertions, whose cogency depends on the inductive evidence that supports them (cf. 1974: 229–30). Take the strongest case, in which the foregoing counterfactual has an unfulfilled antecedent: the particular match was never struck. Perhaps this particular match was destroyed, and hence it will never be struck. Still, we may envisage a *possible world* in which it *was* struck. Would the match, in that possible world, light?

On Mackie’s view, in order to answer this question, we first need to decide what it is reasonable to carry with us “in our baggage for the journey” to this possible world (1974: 201). This decision, he thinks, depends on the *evidence* we have. If the general statement “Matches that are struck get lit” is well-supported and confirmed by evidence in the actual world, then we can put it in our baggage for the journey to the possible world in which the particular match *was* struck. By doing so, we can use it to form a reasonable belief in the counterfactual “If this match had been struck, it would have lit”. As Mackie (1974: 203) notes, the evidence plays a *double role*. It first establishes inductively a generalization. But then, “it continues to operate separately in making it reasonable to assert the counterfactual conditionals which look like an extension of the law into merely possible worlds” (*ibid.*). So it is general propositions (via the evidence we have for them) that carry the weight of counterfactual assertions. If, in the actual world, there is strong evidence for the general proposition “All *Fs* are *Gs*”, then “we feel justified in extending [it]” beyond its observed instances “not only to other actual instances but to merely possible ones” (1974: 55). We base our confidence that “if *x* had been an *F* it would have been a *G*” on the evidence that supports the general proposition. Note that Mackie is no realist about possible worlds. He does not think that they are as real as the actual. Hence, his talk of possible worlds is a mere *façon de parler* (cf. 1974: 199). Still, he thinks that it is convenient to think in terms of possible worlds when we assess counterfactuals.

Mackie’s views on counterfactuals have three important consequences for his views of the meaning of causal statements. First, it *cannot* ground a fully objective distinction between causal sequences

of events and non-causal ones. Indeed, Mackie's theory of the circumstances under which a counterfactual conditional can be asserted is, by and large, epistemic. So whether a sequence will be deemed causal will depend on whether the *evidence* is strong enough to support – via a general statement – the relevant counterfactual. As he says: “the holding of a counterfactual conditional is not a fully objective matter” (1974: 55). Mackie, however, fully accepts this consequence of his views. The sought-after distinction between causal and non-causal sequences, the feature of causation, which, as we saw in section 1.6, Mackie called “*necessity*₁”, turns out to rest on some epistemic fact. In fact, Mackie goes as far as to claim: “Hume's resort to psychology was not wrong in principle but merely premature” (1974: 55). But then, his own account of causation is not radically different from Hume's.¹ Nor from the standard Humean ones, as we shall see in Chapter 5. Besides, although Mackie has aimed to identify an *intrinsic* feature of a causal sequence of events that makes the sequence causal, it is clear that he has failed to do so. Whether a sequence of events will be deemed causal will depend, in his view, on an *extrinsic* feature; namely, on whether there is *evidence* to support the relevant counterfactual conditional. It is for this reason that Mackie goes on (as we shall see in section 4.1) to try to uncover an intrinsic feature of causation, in terms of a *mechanism* that connects the cause and the effect.

The second (related to the first) consequence of Mackie's views is this. Since Mackie's account of the meaning of singular causal statements depends on counterfactuals, and since counterfactuals cannot be deemed to be true or false, it follows that singular causal statements *cannot* be deemed to be true or false (cf. 1974: 54). This is a serious handicap of Mackie's account. One of the important claims of those who defend singular causation is that RVC fails to offer adequate truth conditions to singular causal statements. In this respect, Mackie's theory fails no less.

The third consequence is that Mackie's reliance on general statements – which report regularities – in his account of counterfactuals might make his attempt to offer a genuinely singularist account of the meaning of causal statements a non-starter. If the meaning of singular causal statements is given by counterfactual conditionals, and if these counterfactuals rely on regularities, why is it *not* the case that reference to regularities is part of the *meaning* of singular

causal statements? In reply to this, Mackie (1974: 60) makes a distinction between the *meaning* of a singular causal statement and its *ground*: “the meaning of causal statements is given by the [counterfactual] conditionals, but their grounds may well include the corresponding regularities.” This meaning–ground distinction is based on the thought that singular causal statements do not *imply* any generalizations. One could, of course, contest the claim that singular causal statements do not imply any generalizations. As we saw in section 2.6, Davidson did. But suppose we granted, for the sake of the argument, Mackie’s meaning–ground distinction and conceded that the meaning of singular causal statements is given by the relevant counterfactuals. Still, Mackie would need to explain the meaning of counterfactual conditionals. Now, Mackie’s account of counterfactuals *fails* to offer an intuitively correct meaning to them. Recall that for Mackie the holding of the counterfactual “If x had been an F , then x would have been a G ” depends on the *evidence* we have for a universal generalization “All F s are G s”. Then, the counterfactual will inherit the *uncertainty* of the generalization. The relevant counterfactual would not be “If x had been an F , then x would have been a G ”, but something like this: “*Probably* if x had been an F , then x would have been a G ”. This latter counterfactual is *not* the right sort of counterfactual that characterizes a *causal* sequence of events. Counterfactuals that, intuitively, support a causal sequence are *not* uncertain. If the heating of a gas *caused* it to expand, then the counterfactual “If this gas had been heated, it would have expanded” would be true *simpliciter*, and not just probably so.²

3.1.2 Overdetermination and fragile events

In any case, there is a very central objection to Mackie’s view of causation. It comes from the possibility of causal overdetermination. Here is Mackie’s (1974: 44) own example. A man is about to start a trip across the desert. Unbeknown to him, one of his enemies puts a deadly poison in his can for drinking water. But another of his enemies, not knowing that poison has already been put in the can, makes a hole in its bottom. The man starts his trip, the poisoned water all leaks out before his first drink, and he dies of thirst. What caused the traveller’s death? Surely, it was the hole in

the bottom of the can. Yet Mackie's account fails to deliver this judgement. The puncture in the can was not necessary in the circumstances for the man's death, since even if it was not present, the deadly poison would have caused the death. In cases of causal overdetermination we have two factors, each of which is sufficient to bring about the effect, but none of them is necessary, since even if the one was not present, the other factor would ensure the occurrence of the effect.

Mackie has an obvious retort. If we take the effect "*as it came about*" (1974: 46), it's no longer the case that this *particular* effect would have occurred, even if its particular cause had not. In the example above, the event of the traveller's death was also the event of the traveller's *death from thirst* and not the event of his death by poison. A death by poison is a different event from a death from thirst, and hence, Mackie argues, the latter event was *not* overdetermined. The puncturing of the can was necessary for the traveller's *death from thirst* – and hence it was the cause of death – even if under different circumstances, the poison would be sufficient for the traveller's death.

In offering this response, Mackie has a price to pay. If singular causal statements are such that the effect is taken to be as it really occurred (in its "fullest concrete individuality", as Ducasse (1969: 75) put it), then it turns out that the effect is also necessary-in-the-circumstances for its (temporally prior) cause. Mackie's theory of causation suggests that "*x caused y*" means that in all possible worlds that are like the actual in law and fact, if *x* hadn't happened, then *y* would not have happened. Substitute for *x* the traveller's death from thirst (that is, the effect), as it actually happened, and substitute for *y* the puncturing of his can (that is, the cause). It is easy to see that the traveller's death from thirst (i.e. the effect *x*) was counterfactually necessary-in-the-circumstances for the puncturing of the can (i.e. the cause *y*), and hence that, on Mackie's view, it *caused* the puncturing of the can. For in the possible worlds in which the traveller didn't die *in the very specific way in which he died* in the actual world, his can was not punctured. So Mackie's reliance on "fragile" events allows for *backtracking counterfactuals* and hence for the dependence of the cause on the effect.³ Perhaps, Mackie can avoid this objection by building into causal claims his notion of *causal priority*. He could argue that in the example above,

x (the traveller's death) was *not* causally prior to y (the puncturing of the can), and hence it cannot be the cause of it. I won't discuss here Mackie's theory of causal priority. Suffice it to say that although he thinks that causes are prior to their effects, he does not think that causal priority is assimilated to temporal priority. In Mackie (1974: Ch. 8), he defends an account of causal priority based on the thought that causes are prior to effects in the sense that they are *fixed* (at any time) when their effects are not yet fixed.⁴ But as Beauchamp and Rosenberg (1977: 382–94) point out, Mackie's views of causal priority run into serious trouble.⁵

3.2 Causes as *INUS* conditions

We have already seen, in section 2.2, that Mill offered a sophisticated version of RVC. Although Mackie holds no brief for RVC, he has elaborated further the Millian account in an attempt to formulate RVC in an even more adequate way, before he gives it “a fair trial” (1974: 60). He bases his own account of RVC on two central elements of the Millian view. First, effects are typically brought about by a cluster of factors (conditions), which constitute the whole cause of them. Secondly, effects have, typically, a “plurality of causes” (1974: 61). That is, a certain event-type can be brought about by a number of distinct clusters of factors. Each cluster is sufficient to bring about the effect, but none of them is necessary.

Let us illustrate this by a stock example. A house catches fire and gets burned to the ground. There are a number of clusters of factors that can cause house fires. One cluster includes the occurrence of a short circuit along with the presence of oxygen, the presence of inflammable material in the house, the absence of a sprinkler system and so on. Another cluster includes the presence of an arsonist, the use of petrol, the presence of oxygen etc. Yet another includes the eruption of fire in a neighbouring house and so on. Each cluster is a logical conjunction of single factors. The disjunction of all such clusters (conjunctions) captures the plurality of causes. Each conjunction of factors is sufficient for the fire, but none of them is necessary, since another conjunction of factors can be sufficient for the fire. Supposing that there is a finite number of such clusters, let us symbolize this disjunction of conjunctions as $(ABC \text{ or } DEF \text{ or } GHI)$. The correct version of RVC, Mackie

suggests, should be that whenever an event-type E (e.g. a house fire) occurs, it is preceded by $(ABC \text{ or } DEF \text{ or } GHI)$. In actual instances of an event-type E (e.g. the eruption of fire in a particular house), it will be, typically, the case that only one of the clusters of the disjunction has been instantiated prior to E . But in its full generality RVC should be able to capture the fact that the lawlike connection will be between a disjunctive antecedent of conjunctions of causal factors and an effect. If, apart from being just sufficient for the effect, the disjunctive antecedent is also necessary for the effect, then RVC should take the form:

$$(ABC \text{ or } DEF \text{ or } GHI) \leftrightarrow E.$$

This, Mackie (1974: 62) suggests, should be read as: all $(ABC \text{ or } DEF \text{ or } GHI)$ are followed by E , and all E are preceded by $(ABC \text{ or } DEF \text{ or } GHI)$. How do we get causes out of this version of RVC? How, that is, do we separate causes from conditions on this view? To simplify matters a little, let us suppose that the regularity has the following form:

$$AX \text{ or } Y \leftrightarrow E,$$

where AX and Y are clusters of factors that are minimally sufficient for E . To say that AX is minimally sufficient for E is to say that AX is sufficient for E and that none of its conjuncts (A and X) are redundant: none of them, taken on its own, is sufficient for E ; only both of them in conjunction can bring about E . The conjunction AX , however, is not necessary for E . For E might well occur if Y occurs. Each single factor of AX (e.g. A) is related to E in an important way. It is, as Mackie (*ibid.*) has put it, “an *insufficient* but *non-redundant* part of an *unnecessary* but *sufficient* condition” for E . Using the first letters of the italicized words, Mackie has called such a factor an *inus* condition. Causes, then, are at least *inus* conditions.

Referring again to our stock example, to say that short circuits cause house fires is to say that the short circuit is an *inus* condition for the fire. The short circuit is an insufficient but non-redundant part of an unnecessary but sufficient condition for house fires. It is an insufficient part because it cannot cause the fire on its own (oxygen needs to be present, as well as inflammable material, etc.). It is, nonetheless,

a non-redundant part because, without it, the rest of the conditions (oxygen, presence of inflammable material, etc.) are not sufficient for the fire. It is just a part, and not the whole, of a sufficient condition (which includes oxygen, the presence of inflammable material, etc.), but this whole sufficient condition, whose part is the short circuit, is not necessary, since some other cluster of conditions, for example, an arsonist with petrol, can produce the fire.

It was noted above that causes are *at least inus* conditions. This is not a mere quibble. It is consistent with Mackie's version of RVC that causes can also be either sufficient conditions or necessary conditions, or both. A causal regularity can have any of the following forms:

- i. $A \leftrightarrow E$
- ii. $AX \leftrightarrow E$
- iii. $A \text{ or } Y \leftrightarrow E$
- iv. $AX \text{ or } Y \leftrightarrow E$.

Of these forms, only (iv) has A to be an *inus* condition for E . According to (i), A is a sufficient and necessary condition for E ; according to (ii) A is an insufficient but necessary part of a sufficient and necessary condition for E ; and according to (iii) A is a sufficient but not necessary condition for E . In any case, Mackie (1974: 71) says, to call a factor causal is to say that it is either an *inus* condition (form (iv) above) or better than an *inus* condition (forms (i)–(iii) above) for an event-type E . Mackie's improved version of RVC entails that the generic claim to which this view is committed is this:

For some X and for some Y (which may, however, be null), all $(AX \text{ or } Y)$ are E , and all E are $(AX \text{ or } Y)$ (cf. 1974, 71).

An important merit of this view is that it accommodates the thought that if there are any regularities in the world, they are complex regularities. Besides, it allows for what Mackie calls "*elliptical* or *gappy* universal propositions" (1974: 66). Suppose that the regularities in the world are of the complex disjunction-of-conjunctions type that Mackie (and Mill) have envisaged. Our knowledge of these regularities will be, for the most part, gappy or elliptical. We know, for instance, a number of factors that cause death in

human beings, but we don't know them all – and perhaps, we shall never know them all. Nor do we know what other factors act with the known ones to cause death. The known factors can be seen as *inus* conditions for death. This partial knowledge can be captured by “*elliptical* or *gappy* universal propositions” of the form: All ($A \dots B \dots$ or $D \dots \text{not-}C$ or $K \dots M$) are E and all E are ($A \dots B \dots$ or $D \dots \text{not-}C$ or $K \dots M$). Given what we said above, it is clear that the known factors (A, B, D , etc.) in the above elliptical generalizations are *inus* conditions. If Mackie's version of RVC is correct, there must be a full universal proposition that completes the gappy or elliptical one. The more the latter is filled in, the more we know about the full complex regularity; and hence about, say, the causes of death in human beings. Furthermore, if Mackie's version of RVC is correct, we can see how we can engage in causal inference. Take a regularity of the form AX or $Y \leftrightarrow E$. Suppose that an instance of E has occurred. If we also happen to know that, in this particular instance, Y has not occurred, we can infer that AX has occurred.⁶

There are two sorts of problems that this sophisticated regularity view faces. The *first* is that it fails to distinguish between genuine causes and mere joint effects of a common cause. Take Mackie's own example (1974: 84). Workers in London and workers in Manchester stop work at 5pm after the sounding of the hooters in the factories of London and Manchester respectively. Clearly, both events are joint affects of a common cause; namely, that workers stop at 5pm with the sound of a hooter. Yet it can be shown that the sounding of the hooters in Manchester can be an *inus* condition, and hence the cause, of the workers stopping work in London. Take A to be the sounding of the hooters in Manchester and E to be the stopping of work in London. Call X the conjunction of the presence of whatever factors ensure that Londoners stop work at 5pm (e.g. that some automatic devices set off the hooters) and the absence of whatever factors would make the Manchester hooters sound if it wasn't 5pm. Call Y the conjunction of another set of factors that would be sufficient to make Londoners stop work, in case, for instance, there was a power cut and the London hooters couldn't sound. So we have a complex regularity of the form $(AX \text{ or } Y) \leftrightarrow E$. It's easy to see that A (the sounding of the hooters in Manchester) is an *inus* condition for E . A is clearly non-redundant in the conjunction AX , for it alone ensures that it should be 5pm when the hooters

sound. Yet *A* (the sounding of the hooters in Manchester) is not the cause of *E*, that is, of Londoners' stopping work at 5pm. So an event can be an *inus* condition for another event, without being its cause.

Mackie does think that counter-examples such as the above do some serious damage to the thought that all there is to causation is regularity (even of the complex form that the *inus*-condition approach suggests). Here too, he takes it that the remedy would be to introduce some notion of causal priority. Two joint effects of a common cause might be such that one is, from a technical perspective, an *inus* condition for the other. But, Mackie (1974: 85) suggests, two effects of a common cause are not causally prior to each other: they both get fixed after the cause gets fixed. As I noted in section 3.1.2, however, Mackie's theory of causal priority faces some insurmountable problems. Incidentally, there is another reason why Mackie rejects RVC; namely, that it fails to offer a robust account of the necessity that is supposed to characterize causal sequence. Since Mackie's alternative is couched in mechanistic terms, we shall discuss it in some detail in section 4.1.

The second problem faced by the *inus*-condition approach has been diagnosed by Jaegwon Kim (1971): the characterization of causes as *inus* conditions is description- (or language-)dependent. So what causes what will end up being dependent on how the relevant situation is being described (1971: 434). Take the case in which there is a complex regularity of the form $(A \text{ or } (\text{not-}A \text{ and } B)) \leftrightarrow E$. *B* is an *inus* condition for *E*. But $(A \text{ or } (\text{not-}A \text{ and } B))$ is logically equivalent to $(B \text{ or } (\text{not-}B \text{ and } A))$. It is natural to think that both $(A \text{ or } (\text{not-}A \text{ and } B))$ and $(B \text{ or } (\text{not-}B \text{ and } A))$ describe the same event. If we didn't think that, we would have to accept the implausible view that a logically equivalent description of an event can somehow change the event into a different one. Since both $(A \text{ or } (\text{not-}A \text{ and } B))$ and $(B \text{ or } (\text{not-}B \text{ and } A))$ are logically equivalent, we can rewrite the complex regularity of the form $(A \text{ or } (\text{not-}A \text{ and } B)) \leftrightarrow E$ as $(B \text{ or } (\text{not-}B \text{ and } A)) \leftrightarrow E$. But although these two expressions are notational variants, according to the first it is *B* that is an *inus* condition for *E*, whereas according to the second, it is *not-B*. Perhaps this objection is not so damaging as it first seems. As Kim suggests, it does point to the need to have a more rigorous account of what kinds of events can be the relata of causal relations. A mere syntactical characterization of events (especially when it comes to

talk of disjunctive events, or conjunctive events, or complex ones) is bound to lead to failures such as the one Kim's counter-example brought to light.

3.3 Counterfactual dependence

Lewis aims to analyse causation in terms of non-causal facts, but these facts involve relations of counterfactual dependence between events, which are themselves analysed in terms of relations among possible worlds. Regularities do enter the picture, but as a means to ground claims about the relations of similarity among possible worlds. So, despite the fact that regularities (in the form of laws of nature) play an essential role in Lewis's account of the truth conditions of counterfactuals, his account of causation is not a version of RVC.⁷ Lewis's counterfactual theory can be seen as an attempt to improve on Mackie's account. Unlike Mackie, Lewis (1973) puts forward an *objectivist* theory of counterfactuals, based on possible-worlds semantics. Let me start with an outline of this theory.

3.3.1 Possible worlds and miracles

A possible world is a way the world might be, or might have been. For instance, it is possible that gold is not yellow, or that planets describe circular orbits, or that birds do not fly, or that beer doesn't need yeast to brew, and so on. All these situations, which are merely possible from the point of view of the actual world, really *occur* in some possible worlds. So to say that it is possible that gold is not yellow is to say that *there is a possible world in which gold is not yellow*. The totality of facts in the actual world, the way the world *actually* is, comprise one among the many ways the world could be. So the actual world is one among the many possible worlds. Lewisian possible worlds are no less real than the actual. And there are plenty of them. If two possible worlds differ in some facts, or in some laws, then they are different worlds. But then it seems possible to rank worlds according to how *similar* they are. To simplify matters, let's call @ the actual world. Initially, Lewis introduced a *primitive* notion of "comparative overall similarity" according to which "we may say that one world is closer to actuality than another if the first resembles our actual world more than the second

does, taking account of all the respects of similarity and difference and balancing them off against one another” (1986c: 163).

Given this account of possible worlds, take a counterfactual conditional $p \Box \rightarrow q$. For instance, take the counterfactual that if this pen had been left unsupported (p), it would have fallen to the floor (q). Neither p nor q are true of the actual world. The pen was never removed from the table, and it didn’t fall to the floor. Take those possible worlds in which p is true. Call them p -worlds. Then, the counterfactual $p \Box \rightarrow q$ is true (in @) iff the p -worlds in which q is true (i.e. the pen is left unsupported and falls to the floor) are closer to the actual world @ than the p -worlds in which q is false (i.e. the pen is left unsupported but does not fall to the ground, e.g. it stays still in mid-air). As Lewis put it: “a counterfactual . . . is true iff it takes less of a departure from actuality to make the consequent true along with the antecedent than it does to make the antecedent true without the consequent” (1986c: 164). How do we get causation out of all this? We need a preliminary step and three more.

- The *preliminary* step is to note that the relata of causal relations are events. Lewis is interested in “causation in particular cases” (1986c: 161), and he therefore takes events to be event-tokens. Counterfactual conditionals relate propositions, but this is not a problem since to every event e , there corresponds the proposition $O(e)$, which says that e occurs.
- The first step then is to define a notion of *counterfactual dependence* between (families of) propositions. Take the simplest case of two distinct events c and e . There are two families of propositions; namely, $\{O(c), \text{not-}O(c)\}$ and $\{O(e), \text{not-}O(e)\}$. The family $\{O(e), \text{not-}O(e)\}$ counterfactually depends on the family $\{O(c), \text{not-}O(c)\}$ iff the following two counterfactuals hold:

$$\begin{aligned} O(c) \Box \rightarrow O(e) \\ \text{not-}O(c) \Box \rightarrow \text{not-}O(e). \end{aligned}$$

That is, the family $\{O(e), \text{not-}O(e)\}$ counterfactually depends on the family $\{O(c), \text{not-}O(c)\}$ iff it is the case that if c had occurred, then e would have occurred, and that if c hadn’t occurred, then e wouldn’t have occurred.

- The second step is to define *causal dependence* in terms of

counterfactual dependence. So

e causally depends on c iff the family $\{O(e), \text{not-}O(e)\}$ counterfactually depends on the family $\{O(c), \text{not-}O(c)\}$.

- The third step is to get causation out of causal dependence. As Lewis notes, “causal dependence among actual events implies causation” (1986c: 167). So, causal dependence between actual events is sufficient for causation. If two events c and e are actual, and e is counterfactually dependent on c , then c is the cause of e . For instance, let c be the actual short circuit and e be the actual fire. If it is the case that if c hadn’t occurred, then e wouldn’t have occurred, then the short circuit is the cause of the fire. But causal dependence is not necessary for causation. That is, causation does not imply causal dependence. The reason is this. Let e' be an effect of the fire e , for example, that the owner of the burnt house got some insurance money. If c causes e and e causes e' , then c causes e' . That is, causation is *transitive*. Causal dependence, however, is not necessarily transitive, since counterfactual dependence is not. The owner’s insurance compensation (e') is counterfactually dependent on the house having a fire (e) which is, in turn, counterfactually dependent on the short circuit (c). Yet e' is not counterfactually dependent on c . The owner would have got the insurance compensation (e') even if the short circuit (c) had not occurred, assuming that the fire was caused in some other way. So, to complete his analysis of causation in terms of causal dependence, Lewis introduces a way to enforce the transitivity of causal dependence: the sequence of events must form a *causal chain*. A sequence of events $\langle c, e, e', \dots \rangle$ is a chain of causal dependence iff e causally depends on c , e' causally depends on e , and so on.

After these three steps, Lewis (1986c: 167) can argue that “one event is a cause of another iff there exists a causal chain leading from the first to the second”. A possible cause of concern with Lewis’s theory stems from his account of counterfactuals. For many philosophers, there is only one world; namely, the actual. So they think that Lewis’s reification of possible worlds inflates our metaphysical commitments unnecessarily, and is therefore, otiose. In any case, one can wonder what criteria we should employ in

order to rank possible worlds *vis-à-vis* their similarity to the actual world. An unrefined notion of overall similarity will leave a lot of leeway in our judgements as to what counterfactuals come out true. Take a standard example:

If the president had pressed the button, a nuclear war would have ensued (cf. Horwich 1987: 172).

We want to say that this counterfactual is true. But on Lewis's account, it will be false. For a possible world W_1 in which the president did press the button and a nuclear war did erupt is more distant from (because more dissimilar to) actuality than a world W_2 in which the president did press the button but, somehow, a nuclear war did not follow.

In an attempt to dispel these worries, as well as to offer a more informative account of the notion of comparative similarity among possible worlds, Lewis (1986d) goes on to introduce some criteria that are involved in the ranking of worlds (put in order of importance).

- Avoid big, widespread violations of the laws of nature of the actual world (very important).
- Maximise the spatiotemporal perfect match of particular matters of fact.
- Avoid small, localised violations of the laws of nature of the actual world.
- Secure approximate similarity of particular matters of fact (not at all important).

So a world W_1 which has the same laws of nature as the actual world @ is closer to @ than a world W_2 which has different laws. But in so far as there is exact similarity of particular facts in large spatiotemporal regions between @ and a world W_3 , Lewis allows that W_3 is close to @ even if some of the laws that hold in @ are violated in W_3 . As Lewis (1986c: 164) has put it: "similarities of laws are weighty. Weighty, but not sacred." Armed with this, Lewis (1986d: 63–5) argues that his criteria can block the objection based on the president-counterfactual. For, they *disallow* big violations of laws of nature (that is, big *miracles*). To see how the foregoing counterfactual is indeed true, Lewis invites us to consider the

following. Take a world W_1 in which nothing extraordinary happened between the president's pressing the button and the activation of the nuclear missiles. In W_1 the nuclear war did erupt. Take, now, a world W_2 in which the president did press the button but the nuclear war did *not* follow. For this eventuality to happen, many miracles would need to take place (or, to put it in a different way, a really *big* miracle would have to occur). For, all the many and tiny traces of the button pushing would have to be wiped out. Hence, appearances to the contrary, W_2 would be more distant from (because more dissimilar to) actuality @ than W_1 . The *big* violation of laws of nature in W_2 is outweighed by the maximization of the perfect spatiotemporal match of particular matters of fact between W_1 and @. So, with the help of the refined criteria of similarity among possible worlds, the president-counterfactual comes out true. This is all fine. But, as Horwich (1987: 171–2) has noted, it makes Lewis's theory psychologically implausible. The criteria are so tailored that the right counterfactuals come out true. But they have little to do with our pretheoretical understanding of judgements of similarity.

3.3.2 Pre-emption

Lewis's account also faces the problem that plagued Mackie's counterfactual analysis of causation; namely, causal overdetermination. Cases such as the desert trip discussed in section 3.1.2 suggest that there can be causation without counterfactual dependence. Here is another example. A man is shot in the head simultaneously by two people, who act independently of each other. The man dies and we want to say that each of the shots was a cause of his death. Yet the death is counterfactually dependent on neither of them: if one of them hadn't occurred, the other would have sufficed to cause the death. So it is an implausible consequence of Lewis's views that neither of the two shots was the cause of the death. Lewis (1986c: 194) replies that cases of overdetermination do not pose a threat to his analysis because it is unclear how to apply causal terminology to such cases. Suppose that Lewis is right in his diagnosis. It is still a problem for his own theory that, in cases of overdetermination, it yields a definite *negative* answer to the question of whether each of the shots caused the death (cf. Horwich 1987: 169).

Lewis claims that his account of causation can deal satisfactorily with cases of pre-emption. A pre-empted cause c' is an event that would have led to a certain effect e , but it is such that its occurrence is blocked (or pre-empted) by the occurrence of another event c , which nonetheless causes e . Suppose that two men Mr White and Mr Pink, independently of each other, are set on killing Mr Smith. Unbeknown to them, they make very similar arrangements. They are positioned close to Mr Smith at about the same time and they have him on target. Mr White fires his shotgun; the bullet takes its course and strikes Mr Smith in the head. Mr Smith dies soon after. Mr Pink was ready to fire his shotgun, and had he fired it, given his position, his shooting skills, and so on, the bullet would have also struck Mr Smith in the head, leading to his death. But Mr White's shot scares off Mr Pink, who then flees the scene. Mr Pink's shot is a potential alternative cause of Mr Smith's death: it was pre-empted by Mr White's shot, but had it not been pre-empted, it would have caused Mr Smith's death.

Cases such as these seem to reinforce the claim that there is causation without counterfactual dependence. Mr Smith's death is not counterfactually dependent on Mr White's shot, since if Mr White hadn't fired his shot, Mr Pink would have fired his own, and Mr Smith would have died anyway. In reply to such cases, Lewis (1986c: 171–2) appeals to his notion of causal chains. Mr Smith's death is counterfactually dependent neither on Mr White's shot nor on Mr Pink's. However, he claims, we can still say that it was Mr White's shot that caused Mr Smith's death, since there is *a causal chain of actual events* that connects Mr White's shot and Mr Smith's death, whereas there is no such chain of events that forms a causal chain between Mr Smith's death and Mr Pink's shot. Take an intermediate event d (e.g. that the bullet passed in between two trees) between Mr White's shot c and Mr Smith's death e . Then e is causally dependent on d and d on c . And $\langle c, d, e \rangle$ is a *causal chain* and it is in virtue of this chain that c caused e .

One, however, may doubt that the effect e counterfactually depends on the intermediate event d . One might reason as follows: if d had been absent, c would also have been absent, since d is there because c caused it to be; but then c' (the pre-empted cause) would have occurred and caused e . In order to deal with this worry, Lewis appeals to the *direction of causal dependence*. In effect, he denies

that the absence of the intermediate event *d* would have made any difference to the occurrence of *c*. So he denies the following counterfactual: if *d* had been absent, *c* would also have been absent. This would be a *backtracking counterfactual*: it would make a temporally prior event be counterfactually dependent on a temporally posterior event. Lewis (1986d) has built into his theory of counterfactual dependence (and hence of causation) an account of the direction of this dependence, which disallows backtracking counterfactuals. In broad outline, his suggestion is this. There is an asymmetry between the past and the future: the former is *fixed*, whereas the latter is *open*. This asymmetry is accounted for in terms of the asymmetry of counterfactual dependence. The past is “counterfactually independent of the present”, since it would remain the same whatever we did now. But the future is not. It depends counterfactually on the present: on what we do now (1986d: 38). Lewis argues that this asymmetry of counterfactual dependence is the result of a *contingent* fact, namely, that every event is excessively overdetermined by subsequent events, but it is scarcely overdetermined by its history. Take the sinking of the *Titanic*. It had some past determinants, that is, some events that led to the sinking (e.g. the collision with the iceberg), but a great number of future determinants, that is, *traces* of this event (the dead bodies, the survivors, the shipwreck and so on, including some traces “so minute . . . that no human detective could read”). There is no “lawful” way in which the combination of all these traces could be there in the absence of the (earlier) event that produced them (cf. Lewis 1986d: 50).⁸

There are two lines of response to Lewis. The first is to challenge the alleged difference between normal and backtracking counterfactuals (cf. Horwich 1987: 161–3, Hausman 1998: 124). It seems that there are circumstances in which we make true backtracking counterfactuals. Take, for instance, a very cautious acrobat who jumps only if there is a net underneath. We could then say, truly, that *had the acrobat jumped, there would have been a safety net*. Here, the antecedent of the counterfactual (the jumping) is temporally posterior to its consequent (the installation of a safety net).

The second line of response is to argue that the problem of preemption remains intact if we consider cases where *c* causes *e* *directly*, that is, cases in which there is no chain of intervening events that lead from *c* to *e*. Here is Scriven’s apposite example:

Suppose we hit an unstable trans-uranic atom with a hopped-up proton in an accelerator and kick an electron out of the outer ring. Suppose that the atom would soon have emitted that electron spontaneously in the natural decay process if we hadn't intervened. (1975: 8)

Indeed, although it is generally acknowledged that Lewis's theory may be able to deal successfully with cases of *early* pre-emption (such as the case of Mr White's shot and Mr Smith's death discussed above), there are cases of *late* pre-emption, which remain recalcitrant. These are cases in which the pre-empting cause *c* brings about the effect *e*, while the pre-empted cause *c'* is still on its way, and runs into completion only *after* the effect *e* has been brought about. Here is Lewis's own example:

Billy and Suzy throw rocks at a bottle. Suzy throws first, or maybe she throws harder. Her rock arrives first. The bottle shatters. When Billy's rock gets where the bottle used to be, there is nothing there but flying shards of glass. Without Suzy's throw, the impact of Billy's rock on the intact bottle would have been one of the final steps in the causal chain from Billy's throw to the shattering of the bottle. But thanks to Suzy's pre-empting throw, that impact never happens. (2000: 184)

In cases of *late* pre-emption, there is no way in which Lewis's reply to the early pre-emption case can be offered. For there is no counterfactual dependence between *c* (Suzy's throw) and *e* (the smashing of the bottle), since *e* would have been brought about by *c'* (Billy's throw). Nor is there a chain of counterfactual dependence of *e* on *c*. For any event *c_i* between *c* and *e*, it can be shown that *e* does not counterfactually depend on *c_i*, since *e* would have happened because of *c'*. Lewisian scholars, including Lewis himself, have tried to account for the cases of late pre-emption.⁹ But, whatever one makes of these attempts, there is a recently noted case that takes Lewis's view to task: *trumping pre-emption*. According to Schaffer (2000), trumping occurs in the following situation: two causes *c* and *c'* can produce an effect *e*, but, under the circumstances of their occurrence, only one of them is the exclusive factor for bringing about *e*. Here is Schaffer's example: "the major and the

sergeant stand before the corporal . . . [they] both shout ‘Charge!’ at the same time, and the corporal decides to charge. Orders from higher-ranking soldiers trump those of lower rank” (2000: 175). “I hope you agree”, says Schaffer, “that the major’s order, and not the sergeant’s, causes the corporal’s decision to charge.” That’s entirely correct, yet it’s not delivered by Lewis’s theory of causation: the corporal’s decision to charge is *not* counterfactually dependent on the major’s order. It is noteworthy that Lewis and his followers cannot avoid the cases of trumping by the device that avoided cases of standard pre-emption. The trumped process (the sergeant’s order in the above example) remains intact throughout and despite the dominance of the trumping process (the major’s order). Nor can Lewis appeal to a chain of counterfactual dependencies between the effect and the trumping process: these dependencies are broken by the presence of the trumped cause. Indeed, Lewis (2000) suggested a thorough revision of his counterfactual theory of causation in order to accommodate cases of trumping.

Lewis and his followers have made ingenious and lengthy efforts to render the counterfactual theory of causation foolproof. But as Horwich (1987: 171) points out, all this is at the price of making the original, intuitively very plausible, account of causation very complicated and counter-intuitive.¹⁰ To stress a recent point made by Armstrong (1999: 181), the counterfactual theory might be salvaged, but only at the price of becoming *ad hoc* and circular. For the claim that *c* causes *e* does not imply that if *c* hadn’t occurred, then *e* wouldn’t have occurred. If anything, it implies the following: if *c* hadn’t occurred, then *e* wouldn’t have occurred, *unless e* was overdetermined, or *e* came to exist *uncaused*, and so on. Then, it’s clear that analysing causation in terms of counterfactuals would either have to eliminate the unless-clause in an ad hoc way, or (inclusively) to appeal to the causal notions involved in the unless-clause.

So far, we have seen objections to Lewis’s theory, which suggest that counterfactual dependence is not necessary for causation. But is it sufficient? Kim (1973) has presented a number of cases in which there is counterfactual dependence without causation. Consider the following few:

- (a) If yesterday had not been Monday, today would not be Tuesday.

- (b) If my brother had not been born in 1960, he would not have reached the age of 40 in the year 2000.
- (c) If I hadn't written two *ts*, I would not have written *Stathis*.
- (d) If I had not switched on the coffee machine, I would not have drunk this nice cup of coffee.
- (e) If my sister had not given birth at time *t*, I would not have become an uncle at *t*.

None of these cases is a case of causation, but, they all ascertain some relations of counterfactual dependence. Examples (a) and (b) are cases of logical (or conceptual) dependence. Example (c) presents a counterfactual relation between two events, one of which is a constituent of another. Example (d) is a case of two actions, one of which counterfactually depends on the other without being its effect. And example (e) presents a case of non-causal determination. In light of such examples, we can argue that even if there is some counterfactual dependence in causation, there is more than just counterfactual dependence.¹¹

3.4 Counterfactual manipulation

Recently, there have been two notable attempts to offer improved counterfactual analyses of causation, which are similar to, but also interestingly different from, each other. One is by Menzies and Price (Menzies & Price 1993) and the other is by Woodward and Hausman (Woodward 2000; Hausman & Woodward 1999). Both approaches make heavy use of counterfactuals. Both link causation with the notion of *manipulation*. Yet where Menzies and Price aim to ground causation to *human agency*, Woodward and Hausman aim at a more objective account of causation, based on the notion of *intervention*. Let us examine them briefly.

3.4.1 Agency theory

The thought that there is a link between causation and manipulation goes back to von Wright (1973). As he put it: "what confers on observed regularities the character of causal or nomic connections is the possibility of subjecting cause-factors to experimental test by interfering with the 'natural' course of events" (1973: 117). Since, he

thought, manipulation is a distinctively human action, he concluded that “the causal relation [is] dependent upon the concept of human action” (*ibid.*). This conception might sound too anthropomorphic. Don’t we think that there would be causal relations, even if there were no human beings around capable of manipulating the magnitudes related thus? Von Wright was careful to note that the dependence of causation on human action is “epistemological rather than ontological” (*ibid.*). It concerns how causal claims are established and not what causation, ultimately, is. He was quite clear that “causation . . . operates throughout nature independently of agency, also in regions of the world inaccessible to human interference” (*ibid.*). However, he thought, the connection between causation and human action is also “logical”, in the sense that the concept of causation “is connected with features which are peculiar to the concept of action” (*ibid.*). So von Wright’s main point was that the very concept of causation is modelled on the concept of human action, where someone acts freely to bring about something.

Accounts such as the above fell into disrepute because philosophers failed to be persuaded that they eschew anthropomorphism. They conflated, they thought, the epistemology of causation with its metaphysics: how we know that a causal relation holds with what this relation *is*. In reviving the manipulability theory, Menzies and Price (1993), who prefer to call it the “agency theory”, try to turn the charge of anthropomorphism to their benefit, by arguing that the concept of causation can be seen as referring to a *secondary quality* and hence that it can be usefully seen as a parallel to the concept of *colour*. On a popular (dispositional) view of colour, to be, say, *red* is to be disposed to look red to a normal observer under normal conditions. In a similar fashion, they argue, an event *A* is the cause of a distinct event *B* “just in case bringing about the occurrence of *A* would be an effective means by which a free agent could bring about the occurrence of *B*” (1993: 189).¹² As seen in the case of colour, secondary qualities are individuated extrinsically, but this fact does not make them (entirely) subjective entities. If causation is seen as a secondary quality, then, Menzies and Price (1993: 192) argue, it is also specified extrinsically, by being rooted “in the idea of manipulation”. It seems fair to say that the agency theory is a counterfactual theory because causation is not understood in terms of actual manipulations/interventions, but rather in terms of *counterfactual*

ones. So the proper formulation of the theory goes as follows: “a causal relation exists between two events just in case it is true that *if* a free agent *were* present and able, she *could* bring about the first event as a means to bringing about the second” (Menzies & Price 1993: 198).

A natural worry at this point is that the account presently discussed is fraught with circularity, for notions such as *bringing about* are themselves causal. Even the notion of a *free agent* sounds causal, in the sense that agents are free to do X if, at least, they are not *forced* to do *not-X*. In a rather interesting move, Menzies and Price (1993: 195) argue that, far from being circular, their account is *reductive*: it defines causation in terms that “do not depend on any prior acquisition of any causal notion”. Their claim is that from the very fact that, *qua* agents, we *succeed* in bringing about something by acting on something else, we can conclude that we have “direct non-linguistic acquaintance with the concept of bringing about an event”. So they point out that we, *qua* successful agents, are in possession of a non-causal “ostensive definition of the concept of bringing about”. This, it should be noted, might be fine as far as it goes. But it does not go far enough. First, even if Menzies and Price are right in their foregoing suggestion, it does not follow that the concept of causation is *exhausted* by its connection with human agency. Secondly, even if it was the case that the *concept* of causation was exhausted by its connection with the concept of agency, it does not follow that causation, as this is in the world, is connected to human agency. Nor does it follow that, where causation exceeds human agency, the notion of bringing about acquired via successful human action is the same as the notion of bringing about involved in a causal relation that does not (and cannot) involve human action. Briefly put, it seems that successful human action via manipulation and intervention is a *symptom* of a causal relation, but not constitutive of it. For there is causation even where there is no possibility of human intervention.¹³

3.4.2 Causes as levers

Hausman and Woodward’s (1999) and Woodward’s (2000) account is meant to ensure that causation is linked with (counterfactual) manipulation, and, at the same time, to block the thought that causation has a special connection with *human* agency. As they

(1999: 533) put their central thought: “causes are levers that can be used to manipulate their effects”. But, (a) they have a much more general notion of manipulation/intervention, which is by no means restricted to human action; and (b) they clearly see that, the concept of intervention being itself *causal*, an account based on intervention cannot offer a non-circular analysis of causation (cf. 1999: 534–5). On their view “a sufficient condition for *X* to cause *Y* is that interventions on *X* within a certain range are associated with changes in *Y*” (1999: 537).

We shall have the opportunity to examine their notion of *intervention* in some detail in section 7.2. So here, I shall only make a two general points about their approach. First, their own theory is *counterfactual*: what matters is what would happen to a relationship if interventions were to be carried out. In particular, a relationship among some variables (or magnitudes) *X* and *Y* is said to be causal if, were one to intervene to change the value of *X* appropriately, the relationship between *X* and *Y* wouldn’t change *and* the value of *Y* would change. To use a stock example, we can say that the force exerted on a spring causes the change of its length, because there is an invariant relationship (within a certain range of interventions) between the force exerted on the spring and its displacement from its equilibrium position (expressed by Hooke’s law) and because it is true that were one to intervene to change the force, one would change the length too.

It should be noted, however, that the notion of intervention involves two important *idealizations*, which seem to obscure it. The first is brought out by Hausman and Woodward’s (1999: 539) claim that interventions “need not be feasible”. So we are talking about *ideal* interventions, that is, interventions that *could* take place. But then, there seems to be, at least *prima facie*, an issue concerning what interventions are and are not possible and how this is specified.

The other idealization relates to the concept of *modularity* that Hausman and Woodward (1999: 542) introduce. The gist of their idea is this. Suppose that variables *X* and *Y* are parts of a system, whose behaviour is governed by a number of causal laws; one of these is the causal law that connects *X* and *Y*, but there are other causal laws that have *Y* as their effect. Suppose now that there was an intervention *I* on variable *X* which, while changing the value of

variable Y , also changed, or disrupted, independently of the manipulation of X , some other causal law of the system, thereby influencing the value of Y . If this happened, no sound conclusions as to the cause of Y could be drawn. Let's illustrate this by means of an example. Suppose that there is an intervention I on a patient who suffers from severe pain (Y). The intervention consists in giving the patient a painkiller. The pain is alleviated very soon. So the thought may be that the *cause* of the relief from pain (i.e. the change of the value of variable Y) was the fact that the painkiller had the right chemical composition X , which stopped the pain. In particular, the thought may be that the relief was brought about by the fact that there is a causal law connecting the specific chemical composition X of the painkiller and the subsequent relief from pain Y . But can we say this, right away? Not really. For, as is well-known, the intervention I (that is, the taking of the pill) may have an effect on Y (that is, the pain), independently of the fact that the painkiller has the right chemical composition X . It happens all too often that the pain goes away just because the patient who has taken the painkiller *expects* to get better very soon and, as a result of this self-inflicted expectation, gets better. It may even happen that the patient feels better even if the pill was not a painkiller at all, but a *placebo*. So here we have a case in which we cannot really tell whether the change of the value of Y was due to X , because the intervention I can change the value of Y in two ways, either by the chemical composition of the painkiller (X) or by the self-inflicted expectation of getting better.¹⁴

In effect, *modularity* is a requirement on causal systems that ensures that a situation such as the above does *not* occur. Positively put, modularity ensures that intervention on the variables of a causal law do not disrupt, or activate, the other causal laws (if any) of the system. Hausman and Woodward (1999: 549) motivate modularity by claiming that "if two mechanisms are genuinely distinct it ought to be possible (in principle) to interfere with one without changing the other". Yet it's not hard to see that modularity is a very strong idealization, which, as Nancy Cartwright (2000b: Ch. 4) has persuasively argued, breaks down all too often. In fact, it is arguable that modularity holds mainly in randomized double-blind experiments (*aka* clinical trials).¹⁵

A second general point about the Hausman and Woodward approach relates to the following suggestion made by Woodward:

“what matters for whether X causes . . . Y is the ‘intrinsic’ character of the X - Y relationship but the attractiveness of an intervention is precisely that it provides an extrinsic way of picking out or specifying this intrinsic feature” (2000: 204). This might be taken to imply that he allows for a conceptual distinction between causation and invariance-under-interventions: there is an *intrinsic* feature of a relationship in virtue of which it is causal, an *extrinsic* symptom of which is its invariance under interventions. But it seems that Woodward’s position is stronger than this. He seems, in other words, to really mean it that invariance under interventions *specifies* the (intrinsic) feature in virtue of which a relationship is causal. Trying to dismiss the well-known objection that an account such as his conflates the metaphysics of causation with its epistemology, Woodward (2000: 205–6, n.1) notes: “for Y to change under an appropriate intervention X just is what it is for X to cause Y ”. Yet a natural worry that crops up here is that we are still left in the dark as to what exactly it is for X to cause Y , since whatever it is, it is specified in explicitly *causal* terms such as *appropriate intervention*. The problem here is not so much that Woodward ought to have offered a reductive analysis of causation. He is adamant that this cannot be done. Rather, the problem is that even if one were to grant that the otherwise causal notion of intervention, aided by modularity, can be used to *infer* correctly when X causes Y , it might still be the case that X ’s causing Y does *not* just consist in the ground for this inference.¹⁶

Having said all this, it should be stressed that Woodward and Hausman’s approach to causation has cast new light on this troubled concept and, even if it’s not the final word on the matter, it is a very significant step forward. It’s now time, however, to move on to Chapter 4, where we shall examine a prominent attempt to identify the extra element of causation in terms of *causal mechanisms*.

4 Causation and mechanism

Hume couldn't see the link between cause and effect. For his followers, causation, as it is in the objects, just is regular succession. In this chapter, our focus will be some prominent philosophical attempts to show that there is more to causation than regular succession by positing a *mechanism* that links cause and effect. We shall start with Mackie's argument and move on to examine Salmon's and Phil Dowe's theories of causation. In the final section, I shall attempt to offer a conceptual guide to the theories we have discussed in the first part of the book.

4.1 Persistence

Although Mackie has been a critic of RVC, he does not deny that complex regularities are "*part* of causation in the objects" (1974: 194). It is only a part though, since, as he claims, RVC leaves out an important aspect of causation as it is in the objects; namely, *necessity*. As Hume noted, the alleged necessary tie between cause and effect is not observable. But Mackie thinks, not unreasonably, that we may still *hypothesize* that there is such a tie, and then try to form an intelligible theory about what it might consist in. His hypothesis is that the tie consists in a "causal mechanism", that is, "some continuous process connecting the antecedent in an observed . . . regularity with the consequent" (1974: 82). Where Humeans, generally, refrain from accepting anything other than spatio-temporal contiguity between cause and effect, Mackie thinks that mechanisms might well constitute "the long-searched-for link

between individual cause and effect which a pure regularity theory fails, or refuses, to find" (1974: 228–9).

He then goes on to argue that this mechanism consists in the qualitative or structural continuity, or *persistence*, exhibited by certain processes, which can be deemed causal (1974: 218ff). There needn't be some general feature (or structure) that persists in every causal process. What these features are will depend on the details of the actual "laws of working" that exist in nature. For instance, what persists can be "the total energy" of a system, or the "number of particles", or "the mass and energy" of a system (cf. 1974: 217–18). But in so far as something persists in a certain process, this feature can be what connects together the several stages of this process and renders it causal. Mackie illustrates his view by means of the following example. Take Newton's first law, which says that a body retains its state of motion, unless it is acted upon by an external force. Consider a single particle that moves, without interference, in a straight line, according to Newton's first law (see Figure 1).

Mackie claims that

if the particle moves continuously from *A* to *B*, from *B* to *C*, and from *C* to *D*, these being equal distances, in equal times, it is in a very obvious sense keeping doing the same thing. It is not of course logically or mathematically necessary . . . But it well seems expectable that it should go on doing as nearly as possible the same thing . . . if nothing intervenes to make it do anything else . . . [O]n the assumption that there has been and will be no interference, . . . the motion from *A* to *B* has produced the motion from *B* to *C*; but the motion from *B* to *C* is just like the motion from *A* to *B*; so, if like causes produce like effects, the motion from *B* to *C* will produce something like itself, that is the motion from *C* to *D*. (1974: 218)

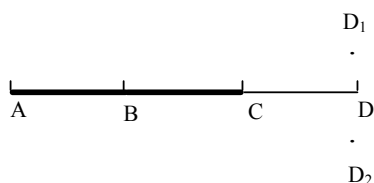


Figure 1

He backs his claim up with the following symmetry consideration. Suppose that after going from A to B and from B to C , the particle goes to D_1 instead of D . This movement

... would have been *prima facie* surprising: since D_1 is placed asymmetrically with respect to the line ABC , we might say that if it were to go to D_1 it might as well go rather to D_2 , similarly placed on the other side of the line; since it cannot do both, it ought not to go to either, but should confine itself to the continuation of its straight path through D . (1974: 220)

For a number of reasons, this particular example might be unfortunate as an illustration of his thesis. Consider the following two. First, Mackie's talk of the motion from A to B "producing" the motion from B to C sounds idle. All that he has managed to establish is the existence of regular succession in the motion of a particle, where equal distances in equal times are succeeded by equal distances in equal times. Secondly, appeal to symmetry considerations shows that the isolation of a pattern of persistence in the process he discusses requires prior *causal* determinations. It is because the particle is not *caused* to go to either D_1 or D_2 that it retains its state of motion and exhibits the persistence that Mackie has identified. Hence, persistence cannot serve as the criterion for the identification of causal sequences, since causal determinations will be required to identify the pattern that persists.¹

Perhaps this particular example shouldn't be taken too seriously. But unless this idea of persistence is filled out, it is not clear that it can carry the weight of necessity – in particular, in a way that, somehow, shows that there is more to necessity than regularity. Take an example that will be used again later on. A moving car casts a shadow on a wall. The shadow moves on along with the car. The shadow is a process, but it is *not* a causal process. To put it crudely, there is nothing inherent in the shadow that keeps it going (but more on why it is not later). However, on Mackie's account of necessity, it might well end up being one. For there are things that persist while the shadow moves. For instance, its shape might well persist over a period of time. Or, the shadow might traverse equal distances in equal times, and so on. Conversely, it might well be the case that there is a causal process between two events, which is not

underpinned by a mechanism. Consider the case of a single radioactive atom that decays after some time t . One might argue that although it can be shown that, given certain statistical laws, this atom has a certain chance to decay in time t , there is no mechanism to ensure that it *will* decay. Its decay (or the lack of it) is, arguably, a purely chancy matter.

Mackie does not say, in any detail, what exactly this structural continuity, or persistence, of a process is. He is, however, quite confident that even when we are faced with a sequence of event-types that are clearly distinct and different (e.g. the striking of a match and the flame), there will be *microscopic* descriptions of these events (e.g. in terms of their molecular and atomic structure) which will show that there is “more [qualitative] continuity and persistence” in the sequence than meets the eye (1974: 221–2). Of course, a process of change cannot be fully reduced to a process of persistence. But Mackie feels that a notion of “partial persistence” should be enough to distinguish between causal sequences of events and non-causal ones. The point remains, however, that talk of persistence (be it partial or whole) remains too vague to substantiate the claim that “qualitative or structural similarity” constitutes “a sort of necessity that may belong to basic laws of working” (1974: 223). Even this last notion (of “laws of working”) is too vague to carry the weight of some non-Humean notion of necessity.²

It should also be noted that Mackie’s account of persistence leaves out one important aspect of causation; namely, causal interaction. When two processes interact (e.g. when a charged particle moves in a gravitational and an electromagnetic field), “it seems inescapable that there should be a law of working which is not just the persistence of anything” (1974: 222).

4.2 Causal processes and mark-transmission

The idea of a causal mechanism is, however, very attractive. Can it be made to work? Salmon has systematically tried to offer an account of the mechanism that links cause and effect. His main aim, as he puts it (1997a: 16), is to “take Hume’s challenge seriously”, that is, “to find a physical connection between cause and effect”. Meeting this aim, Salmon thinks, requires a change in our conceptual machinery. Instead of taking distinct events (or facts) to be the

causal relata, Salmon thinks we should try to characterize directly when a process is causal. So, processes rather than events should be the basic entities in a theory of physical causation.³ Salmon argues that causal processes are “precisely the connections Hume sought, that is, that the relation between cause and effect is a physical connection (although it may not be the *necessary* connection that Hume referred to)” (1997a: 16). In particular, Salmon takes causal processes to be the fundamental element of the mechanistic approach to causation: “they are the mechanisms that propagate structure and transmit causal influence in this dynamic and changing world. . . . they provide the ties among the various spatio-temporal parts of our universe” (1997a: 66).

Where events are happenings that are localized in space and time (e.g. the death of Julius Caesar, or the sinking of the *Titanic*), processes “have much greater temporal duration, and in many cases, much greater spatial extent” (1984: 139). Examples of processes include a light wave travelling from the sun, or less exotically, the movement of a ball. Using the language of the Special Theory of Relativity, we can say that a process is represented by a world line in a Minkowski diagram, whereas an event is represented by a point. Salmon includes in processes material objects (and, in general, physical entities) that persist through time. So he can accommodate within his theory Mackie’s notion of persistence. An important aspect of Salmon’s views is that processes are *continuous*. So a process cannot be represented as a sequence of discrete events. The continuity of the process accounts for, ultimately, the direct link between cause and effect (cf. 1984: 156–7).

Not all processes are causal. In Salmon (1984: 142), borrowing an idea of Reichenbach’s (1956), Salmon characterized as *causal* those (and only those) processes that are capable of transmitting a mark. Consequently, non-causal (or “pseudo”) processes are those (and only those) that cannot transmit a mark. Intuitively, to mark a process is to interact with it so that a *tag* is put on it. A moving white ball (i.e. a process) can be marked by simply painting a red spot on the ball. But it is not enough that the process can be *markable*. The process should be such that, after the mark has been put on it, by means of a single local interaction, the mark gets *transmitted*. Salmon insists on the *transmission* of the mark because without it, there cannot be an adequate characterization of causal processes.

Any process can be marked by means of a single local interaction. Take an example of what Salmon calls “pseudo-processes”; namely, the successive positions occupied by the shadow of a moving car. This is marked when, for instance, the shadow intersects with a pole. The shadow changes shape because of the intersection, but this mark (the changed shape) does not get transmitted beyond the point of the intersection with the pole: the shadow gets restored to its original shape. In order, therefore, to avoid the trivialization of the mark-method, Salmon insists that the mark should be *transmitted* by the process, after the interaction that marked it has taken place (cf. 1984: 142). It was noted above that the interaction that marks a process should be a single local interaction, that is, an interaction “*at a single point* in the process”. This is necessary for the following reason. Any pseudo-process can be made to transmit a mark by means of many suitable interactions. Suppose, for instance, that I distort (that is, I *mark*) the shadow of a moving ball by interpolating my hand between the ball and the shadow. But this *mark* will get *transmitted* only if my hand keeps moving in contact with the shadow, which constitutes further interactions.

We don’t yet have a precise idea of what a mark is. So Salmon (1984: 144) goes on to characterize the mark method a bit more formally. A process, be it causal or not, exhibits “a certain structure”. A causal process is then said to be a process capable of transmitting its own structure. But, Salmon adds, “if a process – a causal process – is transmitting its own structure, then it will be capable of transmitting certain modifications in the structure” (1984: 144). A mark, then, is a modification of the structure of a process. And a process is causal if it is capable of transmitting the modification of its structure that occurs in a single local interaction. It should be noted, however, that Salmon appears to offer two criteria for a process being causal. The first is that it is capable of transmitting its own structure, that is, that it is, in some sense, self-maintaining or self-persisting, or self-determined. This criterion says nothing about marking, unless of course one thinks that the structure that characterizes a process is its own mark. But even so, whether a process is causal will depend, on the first criterion, on whether the process is capable of *transmitting its own structure*. This is important because it makes Salmon’s theory sufficiently different from Mackie’s. For Salmon “persistence of structure” is not enough to characterize a process as causal. As he

notes (1984: 153), even pseudo-processes may exhibit “persistence of structure”. What pseudo-processes cannot do is *transmit* their structure, unless they are under the influence of some “external agency”. The second criterion that Salmon offers is that a process is causal if it is capable of transmitting *modifications* of its structure. This modification is, clearly, a marking of the process. So the second criterion is a genuine marking criterion. However, the two criteria are conceptually distinct (cf. Woodward 1989: 375–6). They are not even necessarily coextensive. For instance, a photon might be rightly deemed as a causal process according to the first criterion, but it seems that it cannot be a causal process on the second criterion, since it admits of no modification of its structure (assuming that it has one).

One issue that crops up here is whether Salmon’s analysis of causation is Humean or not. It will be (broadly) Humean, if it is cast in terms of non-causal notions. But is it? The question is important because one of the criteria of adequacy that Salmon puts forward is that his theory should not violate Hume’s strictures against “the uncritical use of such concepts as ‘power’ and ‘necessary connection’” (1984: 147). So Salmon intends his theory to be (broadly) Humean. Now, the concept of transmission need not be causal. Hence, it can be admitted by Humeans. In fact, as Salmon points out, “ability to transmit a mark can be viewed as a particularly important species of constant conjunction – the sort of thing that Hume recognised as observable and admissible” (1984: 147).⁴

Indeed, Salmon goes on to explicate further why the ability to transmit a mark is not “a mysterious power”. His master thought is that there is no mystery in the view that a mark is transmitted from a point *A* of a process to a subsequent point *B*, if we take on board Russell’s *at-at* theory of motion. According to this theory – which Russell developed as a reply to Zeno’s paradox of the arrow – “to move from *A* to *B* is simply to occupy the intervening points at the intervening instants” (1984: 153). That is, to move from *A* to *B* is to be *at* the intervening points, *at* the intervening times. Salmon (and Russell) argue that this is a *complete* explanation of the motion since there is no additional question (and hence no extra pressure to explain) why (or how) the objects *get* from point *A* to point *B*. Consequently, Salmon (1984: 148) defines mark-transmission (MT) as follows:

- (MT) Let P be a process that, in absence of interactions with other processes, would remain uniform with respect to a characteristic Q , which it would manifest consistently over an interval that includes both of the spacetime points A and B ($A \neq B$). Then, a mark (consisting of a modification of Q into Q'), which has been introduced into process P by means of a single local interaction at point A , is transmitted to point B if P manifests the modification Q' at B and at all stages of the process between A and B without additional interventions.

Before we make two basic comments on this formulation, let me question Salmon's view (expressed a couple of paragraphs above) that mark-transmission can be viewed as a species of constant conjunction. On the one hand, Salmon talks about the *ability* to transmit a mark, and this is clearly distinct from an actual regularity (or constant conjunction). The ability is a capacity or a disposition, and it is essential for Salmon that it is so. For he wants to insist that a process is causal, even if it is not actually marked (cf. 1984: 147). But constant conjunctions are actual. On the other hand, even if we granted (as, I think, we should) that mark-transmission is sufficient for constant conjunction, it is not necessary. The successive stages of the shadow of a car may form a regularity (constant conjunction), but no mark-transmission is involved in it.

Be that as it may, two things are immediately notable in the formulation of MT. First, the first clause of MT strengthens the criteria for a process being causal by introducing a counterfactual characterization; namely, that "the process P would have continued to manifest the characteristic Q if the specific marking interaction had not occurred" (1984: 148). This is a considerable strengthening because the two criteria that we have encountered so far (i.e. transmission of P 's own structure, and transmission of a modification of P 's own structure) make no references to counterfactuals. The strengthening, however, is necessary because it seems that there can be pseudo-processes that satisfy the second clause of MT. Consider an example, attributable to Nancy Cartwright, as Salmon (1984: 148) acknowledges. A rotating beacon is casting a white spot that moves around a circular wall. The spot is marked by interposing a red filter at the wall. This process is not causal because it

violates both of the first two criteria (it does not *transmit* its own structure; nor does it *transmit* the modification to its structure). But we can do the following thing to turn it into a process P that satisfies both of these criteria (in their original form). A tiny fraction before the red filter is put on the wall, we place a red lens in the beacon. As a result of this, the white spot turns red, by means of a single local interaction (i.e. the red filter on the wall), and *remains* so while it moves around the wall (because of the red lens). The moving red spot should now be taken to be a causal process because it would seem that it has transmitted a mark; namely, the mark made by the red filter on the wall. But it has not. The counterfactual conditional introduced in the first clause of MT is meant to block such counterexamples. The marking of the white spot by the red filter on the wall is not counterfactually robust: the spot would have become and remained red regardless of the presence or absence of the red filter on the wall, because of the red lens.

The second notable feature of MT is that it makes extensive reference to the presence and absence of interactions. But isn't the concept of interaction causal? So isn't Salmon's project to offer a Humean account of causation jeopardized? In Salmon (1984: 171), he defines causal interaction (CI) as follows:

CI: Let P_1 and P_2 be two processes that intersect with one another at the spacetime point S , which belongs to the histories of both. Let Q be a characteristic that process P_1 would exhibit throughout an interval if the intersection with P_2 did not occur; let R be a characteristic that process P_2 would exhibit throughout an interval (which includes subintervals on both sides of S in the history of P_2) if the intersection with P_1 did not occur. Then, the intersection of P_1 with P_2 at S constitutes a causal interaction if:

- (1) P_1 exhibits the characteristic Q before S , but it exhibits a modified characteristic Q' throughout an interval immediately following S ; and
- (2) P_2 exhibits the characteristic R before S , but it exhibits modified characteristic R' throughout an interval immediately following S .

An example of CI is the case that Hume so frequently talked about: the collision of two billiard balls. This case satisfies CI, because after the collision, the state of motion of the two balls is modified and the modifications persist beyond the point of collision. But let us take a more careful look at CI.

Note first that the formulation of CI involves, once more, counterfactuals. This is to secure that intersections between pseudo-processes do not count as causal interactions.⁵ Secondly, the actual wording of CI is such that the concept of causal interaction is defined in terms of the geometric (i.e. non-causal) concept of *intersection* of two processes. So it might seem that Salmon can manage to offer an analysis of causation in non-causal terms. But this is not immediately obvious. For CI makes an essential (if implicit) reference to *marks*, and hence to *causal* processes. To see this, it is enough to extract the essence of CI. As Dowe (2000: 71) has stated, CI says: "An interaction is an intersection of two processes where both processes are marked and the mark in each process is transmitted beyond the locus of intersection." Salmon himself has stated that "if two processes intersect in a manner that qualifies as a causal interaction, we may conclude that both processes are causal, for each process has been marked (i.e. modified) in the intersection with the other and each process transmits the mark beyond the point of intersection" (1984: 174).

The problem that arises here is that, as Salmon (1997a: 17, 249) acknowledges, the concept of a mark is itself a *causal* concept. He thinks, however, that the appeal to the non-causal concept of *intersection* is enough to ground his theory in non-causal terms. Endorsing, essentially, Dowe's formulation of CI, he offers the following formulation of his theory (1997a: 250):

- S-I A process is something that displays consistency of characteristics.
- S-II A mark is an alteration to a characteristic that occurs in a single local intersection.
- S-III A mark is transmitted over an interval when it appears at each spacetime point of that interval, in the absence of interactions.
- S-IV A causal interaction is an intersection in which both processes are marked (altered) and the mark in each process is transmitted beyond the locus of the intersection.

- S-V In a causal interaction a mark is introduced into each of the intersecting processes.
- S-VI A causal process is a process that can transmit a mark.

Given this formulation, he is confident that his account is cast in non-causal terms.⁶ But, even if Salmon is right in this, it's not clear that his account in terms, ultimately, of intersections, is strong enough to characterize causal *interactions*. The problem lies with his *S-II*. *S-II* introduces the concept of a mark. A mark is a modification of some kind; "it need not persist" (1997a: 250). Then, *S-IV* and *S-V* characterize causal interactions in terms of persisting marks in *intersecting* processes. But, as Dowe (2000: 72) has pointed out, these will allow for pseudo-processes to count as genuinely causal (i.e. as markable). Take two spotlights that intersect on a screen and then move apart. Suppose that at the moment of intersection a red filter is placed on one of the spots (at its source). The spot then changes colour (i.e. the spot is marked red) and the change persists after the point of intersection. The moving red spot will count as a causal process. Yet there is no causal interaction between the two spots. All there is is an "accidental correlation" between a change in a process (placing the red filter in the spotlight) and its intersection with another process. As Dowe observes, Salmon might be able to block this counter-example by appealing to counterfactuals. For, in the presence of the red filter at the source, the beam would change colour even if the intersection with the other beam had not occurred. But this just reinforces the importance of counterfactuals for Salmon's theory of causation.

A variant of the objection considered thus far has to do with the alleged circularity that Salmon's account faces. Crudely put, the challenge is this. To mark a process is to interact causally with it so that a modification of its structure occurs. And to interact causally with the process is to modify its structure (i.e. to mark it) by means of a process, which is also marked (i.e. by means of a causal process). So the definitions of *causal interaction* and of *marking* seem to be mutually dependent (cf. Dowe 2000: 72). Salmon's modified theory (*S-I* to *S-VI* above) has aimed to block this objection. But he does not seem successful. As Dowe (2000: 73–4) has noted, there is a genuine problem between *S-III* and *S-IV*. *S-III* makes reference to the all-essential mark-transmission, in the absence of further

interactions. But *S-IV*, which characterizes causal interactions, makes reference to mark-transmissions. So, once more, two key concepts seem to be defined in terms of each other. Salmon cannot remove reference to mark-transmission from *S-III*. He takes it to be “fundamental” for the explication of causal processes (1997a: 250). As we have already seen, it is the transmission of a mark that distinguishes causal processes from pseudo-processes. Could Salmon replace the reference to “interactions” in *S-III* with reference to intersections? Could, that is, *S-III* read thus: A mark is transmitted over an interval when it appears at each spacetime point of that interval, in the absence of *intersections*? This would make *S-III* lose its bite. For many genuine causal processes would disqualify from being causal processes, since they may well intersect with lots of pseudo-processes. For instance, the movement of a ball would no longer count as a causal process, since it might well intersect with the shadow of a nearby building.

To sum up, despite Salmon’s intentions to offer a broadly Humean account of causation, it is not clear (to say the least) that he succeeded in this task. Besides, even if we were to treat his account as non-Humean (that is, even if we accepted that his theory should be best seen as a non-reductive account of causation), it seems that it cannot escape from the charge of circularity.⁷

4.3 Marking shadows

Suppose we were to leave aside the problems mentioned in the previous section. The question to ask, then, would be the following: is Salmon’s mark method adequate as a theory of causation? The key element of his theory is the idea of mark-transmission. Is, then, mark-transmission necessary and sufficient for a process being causal? Kitcher (1985) has argued that it is neither. Take the case of a pseudo-process, for example, the shadow of a moving car. This can be permanently marked by a single local interaction. The car crashes on a wall and a huge dent appears on its bonnet. The shadow of the car acquires, and transmits, a permanent mark: it is *the shadow of a crashed car*. So the mark-transmission is not sufficient for a process being causal. Conversely, a process can be causal even if it does *not* transmit a mark. To see how this is possible, consider Salmon’s requirement that a process should remain uniform

with respect to a characteristic Q for some time. This is necessary in order to distinguish a process (be it causal or not) from what Kitcher has aptly called “spatiotemporal junk”. This requirement, however, seems to exclude from being causal many genuine processes that are short-lived, for example, the generation and annihilation of virtual (subatomic) particles (cf. Dowe 2000: 74). Less exotically, Kitcher (1985: 638) invites us to consider the case of a chemical injected into the cytoplasm of a frog zygote. If this is the right sort of chemical, it will be such that it will mark no cell of the young embryo. Yet the frog will display a number of developmental abnormalities. This is a genuine causal process, which is not, however, marked.

A generic problem to which the above counter-examples point is the vagueness of the notion of characteristic Q , which gets either transmitted or modified in a causal process. Salmon could block the first of the counter-examples above by denying, for instance, that the modification of the shadow of the car after the crash is a modification of a genuine characteristic of the shadow. In specific cases, we seem to have a pretty clear idea of what this characteristic might be, for example, the chemical structure of a molecule, or the energy-momentum of a system, or the genetic material of an organism. Once, however, we start thinking about all this in very abstract philosophical terms, it is not obvious that we can say anything other than this characteristic being a *property* of a process. Then again, new problems arise. For at this very abstract level, *any* property of *any* process might well be suitable for offering the markable characteristic of the process. So we seem to be in need of a theory as to which properties are such that their presence or modification marks a causal process. Another generic problem that Salmon’s account faces relates to his view that a causal process is characterized by mark-transmission, *in the absence of further interactions*. As Dowe (2000: 74) has rightly noted, this proviso is unfulfillable. Even in the most idealized cases, there are going to be further interactions present. So Salmon’s account, if taken literally, seems to be in danger of being vacuous.

In the previous section, I noted that Salmon’s theory relies heavily on the truth of certain counterfactual conditionals. As we saw, this holds for both his characterization of MT and his characterization of CI. The counterfactuals seem to play a *double role*. On

the one hand, they secure that a process is causal by making it the case that the process does not just possess an actual uniformity of structure, but also a counterfactual one. This is necessary because pseudo-processes may well exhibit such actual uniformity, but they fail to exhibit a counterfactual one. Since pseudo-processes are such that their uniformity is dependent on some external agency, they would not exhibit this uniformity if the external agency were absent. But causal processes would, since they require no external agency. On the other hand, the counterfactuals secure the conditions under which an interaction is causal. If the marking would have occurred even in the absence of the supposed interaction between two processes, then the interaction is not causal (i.e. it is a mere intersection). Now, Salmon's appeal to counterfactuals has led some philosophers (e.g. Kitcher 1989) to argue that, in the end, Salmon has offered a variant of the counterfactual approach to causation. Such an approach would bring in its tow all the problems that counterfactual analyses face. In particular, it would seem to undermine Salmon's aim to offer an objective analysis of causation, for, as we said in Chapter 3, it is an open issue whether or not there can be a fully objective theory of the truth-conditions of counterfactuals. In any case, Salmon has always been very sceptical about the objective character of counterfactual assertions. So, as he said, it was "with great philosophical regret" that he took counterfactuals on board in his account of causation (cf. 1997a: 18). The question, then, is whether his account could be formulated without appeal to counterfactuals.⁸

The short answer to the above question is: *yes, but . . .* For the mark method has to be abandoned altogether and be replaced by a variant theory, which seems to avoid the need for counterfactuals. The counter-examples mentioned above, as well as the need to avoid counterfactuals, led Salmon to argue that "the capacity to transmit a mark" is not constitutive of a causal process, but rather a "symptom" of its presence (1997a: 253). So causal processes, that is, the "the causal connections that Hume sought, but was unable to find" (1984: 147), should be identified in a different way. We will turn our attention to this issue in section 4.4.

4.4 Conserved quantities

Phil Dowe (2000) has put forward a radical modification of Salmon's theory, which relies on the concept of a *conserved quantity*. As he (2000: 89) put it: "The central idea is that it is the possession of a conserved quantity, rather than the ability to transmit a mark, that makes a process a causal process." Before we explain this idea, it should be noted that Salmon (1997b) has endorsed it, with the further modification that the conserved quantity should be *transmitted* instead of just being possessed. The reason for this switch is, basically, that Dowe's theory frees "the concept of causality from its dependence on counterfactuals" (Salmon 1997a: 260). So a lot will depend on whether or not the Conserved Quantity (CQ) theory can dispense with counterfactuals.

The CQ theory rests on the following two propositions:

CQ-D1 A *causal process* is a world line of an object that possesses a conserved quantity.

CQ-D2 A *causal interaction* is an intersection of world lines that involves exchange of a conserved quantity.

A conserved quantity is "any quantity that is governed by a conservation law" (Dowe 2000: 91). Examples of such quantities are mass, energy, linear momentum and charge. In any case, since Dowe aims to offer an *empirical* theory of physical causation, and in particular a theory of causation as it is in the actual world, he says that we should look to our best scientific theories for what these conserved quantities are. Although conservation laws are involved in the identification of the conserved quantities, and hence in what processes count as causal, Dowe insists that his account of causation is *singularist*. This is so because whether a process is causal will depend "only on local facts about the process" and in particular on whether the process possesses a conserved quantity (2000: 96). So the causal process need not instantiate a regularity. Salmon, who also endorsed a version of the CQ theory, has a similar view. He thinks that this theory "does not require *laws* of conservation" (1997a: 20). All it requires is the truth of the relevant conservation statements, irrespective of whether or not they express a lawful regularity or not (cf. 1997a: 259).

There have been attempts before Dowe to analyse causation in terms of some conserved quantities. David Fair (1979), for instance, suggested that causation amounts to “energy-momentum transference” between physical events or entities. But this view did not find fertile ground among philosophers, partly because it was considered too physicalistic and partly because it seemed to face insuperable problems. Consider just two of those. First, it leaves out some cases of causation, where there is no transference of energy-momentum. These are cases of persistence, such as the ones discussed by Mackie in section 4.1. A body that moves in rectilinear motion without any external forces is such that there is no transfer of energy-momentum to and from it. Yet it seems at least to make sense to say that the cause of its motion is its inertia, that is, that its earlier states cause its later ones (cf. Dowe 2000: 52). In fact, advocates of the persistence view of causation argue that processes where nothing gets transferred from their earlier stages to their later ones, that is, processes with unchanging persisting properties, are among the most fundamental causal processes. As Ehring (1997: 122) put it: “Unchange as well as change falls within the causal structure of the world”. Secondly, transference accounts seem to leave unexplained how energy-momentum gets transferred from one object to another. In particular, they leave unexplained how the quantity that gets transferred remains identical over time (cf. Dowe 2000: 55–9). Dowe’s account is meant to be an improvement over earlier physicalist approaches because it does not involve the notion of transference.

But is Dowe’s “*possession* of a conserved quantity” sufficient for causation? Let us first note that for Dowe *possesses* is to be understood as *instantiates* (2000: 92). So to say that object *x* possesses momentum *q* is to say that the property *having-momentum-q* is instantiated by this object. In particular, Dowe (*ibid.*) does not require either that the conserved quantity be transmitted or that it be kept constant. Given this, it turns out that Dowe can effectively deal with the counter-examples that plagued Salmon’s mark-theory. For instance, it is easy to see that typical cases of pseudo-processes (such as shadows) will be considered as pseudo-processes by Dowe’s theory, since they possess no conserved quantities.

As Christopher Hitchcock (1995) has pointed out, there seem to be pseudo-processes that can be said to possess some conserved

quantity. Take, for instance, a moving shadow that is cast on a metal plate, which has a uniform charge density on its surface. The shadow will then possess a quantity of electric charge, which is a conserved quantity. How can Dowe reply to this? His answer (2000: 98) is that in cases such as the above, it is not the shadow (or the pseudo-process, in general) that possesses the conserved quantity. Rather, it is the charged plate that does. But there is a modification of the counter-example that Dowe cannot meet so easily. Imagine, Salmon (1997b: 472) says, that we take the aggregate of the patches of the charged plate that are in shadow, while, and as long as, they are in shadow. This aggregate of patches that are sequentially in shadow, taken only for the time they are in shadow, involves some uniformity and displays spatiotemporal continuity. Besides, it possesses a conserved quantity; namely, charge. Yet it is *not* a causal process. So Dowe seems wrong to say that possessing a conserved quantity is sufficient for a process being causal. How does Dowe reply to this? He claims (2000: 99) that the foregoing *aggregate* is not a proper process; it is a gerrymandered one. His main argument is that it is a necessary condition for a process being causal that it exhibits identity through time. Aggregates such as the above do not display such identity over time. Take an example he offers (2000: 99). A putative object x might be defined as follows:

- for $t_1 \leq t < t_2$; x is the coin in my pocket
- for $t_2 \leq t < t_3$; x is the red pen on my desk
- for $t_3 \leq t < t_4$; x is my watch.

x , Dowe says, occupies a determinate spacetime region, and at any time in the interval t_1 to t_4 , it possesses conserved quantities (e.g. momentum). Yet it cannot be a causal process, for it cannot be an object. It cannot be an object because it “fails to display identity over time” (2000: 100). It simply consists of “a collection of different objects at different times” (*ibid.*).

The consequence of all this is that Dowe’s CQ theory will be very sensitive to how processes are individuated and, in particular, to how the notion of identity-through-time should be analysed. For instance, on pain of circularity, Dowe cannot analyse the notion of identity-through-time in terms of the (currently popular) causal

theory of identity. To cut a long story short, Dowe's (2000: 107) view is that the notion of identity of an object is a "primitive" concept of his theory. This is deeply unsatisfactory, however. Dowe's admission amounts to the claim that, ultimately, which processes will count as causal will also be a primitive concept of his theory.

It was noted above that Salmon has endorsed the basics of Dowe's CQ theory. Their disagreement is precisely over the issue of whether the conserved quantity should be simply possessed by the causal process, or whether it should be transmitted by the process. Salmon (1997b) takes the latter view. His main argument for this is that the concept of transmission can ground the difference between causal and non-causal processes. Consider the foregoing counter-example with the charged plate. Salmon's own reply to this is that the aggregate of patches of the charged plate that are sequentially in shadow is not a causal process because the conserved quantity (i.e. charge) is not *transmitted* among the patches. So Salmon's version of the CQ theory consists in the following three definitions (cf. 1997b).

- CQ-S1 A causal process is a world line of an object that transmits a non-zero amount of a conserved quantity at each moment of its history (each spacetime point of its trajectory).
- CQ-S2 A causal interaction is an intersection of world lines that involves exchange of a conserved quantity.
- CQ-S3 A process transmits a conserved quantity between *A* and *B* ($A \neq B$) if and only if it possesses [a fixed amount of] this quantity at *A* and at *B* and at every stage of the process between *A* and *B* without any interactions in the open interval (*A*, *B*) that involve an exchange of that particular conserved quantity.

CQ-S1 differs from Dowe's CQ-D1 in two respects. First, it insists on the transmission of the conserved quantity. Secondly, it requires that the process transmits a non-zero amount of a conserved quantity. This is because Salmon (1997a: 256) wants to block the following type of counter-example. It can be argued that a pseudo-process (e.g. a shadow) possesses (and transmits) a *zero* amount of a conserved quantity (e.g. charge). If it weren't for Salmon CQ-S1, this pseudo-

process would qualify as causal. Dowe's reaction to this is that there is no need to require the amount of the conserved quantity to be non-zero because, as he says, shadows and other typical pseudo-processes "are not the type of objects to which conserved quantities may be ascribed" (2000: 118). CQ-S2 is the same as Dowe's CQ-D2. CQ-S3 is Salmon's addition to the CQ theory. It is meant to explain the concept of transmission, based on Salmon's earlier *at-at* theory (see section 4.2). Dowe denies that the concept of transmission is necessary for the CQ theory. But what really hangs on this? Why does Salmon insist on transmission, while Dowe doesn't? I am not sure what really is at stake here. It seems that the two versions of the CQ theory are (almost) equivalent. For, it seems that the following conceptual equation holds:

transmission of a conserved quantity P = possession of a
conserved quantity P + identity over time of the object
that possesses P .

Salmon holds the left-hand side of this equation, whereas Dowe holds the right-hand side. Salmon thinks that Dowe's appeal to an unanalysed concept of identity-over-time is problematic (1997b: 469). Dowe, on the other hand, thinks that Salmon's concept of transmission sneaks in a preferred direction in causal processes. If the transmission of the conserved quantity is *from* spacetime point A to spacetime point B , then, Dowe argues, this order will offer direction to causation: the asymmetric relation between cause and effect. For reasons that need not concern us here, however, Dowe takes it to be an advantage of his own version of the CQ theory that it is "non-committal on the question of causation's direction" (2000: 110). It should be noted, though, that Salmon's formulation of CQ-S3 is also noncommittal on the question of the direction of causation.⁹

Be that as it may, both Dowe and Salmon commend the Conserved Quantity theory because it manages to avoid any reference to *counterfactuals*. As was noted in the beginning of the section, this is a crucial issue, so we need to examine whether it can indeed avoid any reliance on counterfactuals. On the face of it, it can. Both versions of the theory characterize a process as causal in terms of the *actual* possession or transmission of a conserved quantity. But if we dig deeper some, at least *prima facie*, problems might appear.

A *first* worry is based on Dowe's reaction to CQ-S1 above. Shadows with zero quantity of charge are still pseudo-processes, since, Dowe says, they are not the type of object "to which conserved quantities may be applied" (2000: 118). Yet particles at rest, which have zero momentum, are causal processes because they are the type of object to which conserved quantities (even of value zero) may be applied. But how is this difference to be grounded? I think that an appeal to modalities is inescapable. Take the case of a particle at rest, that is, a particle with zero momentum. Suppose that this particle does not enter into any interactions, so that its state of motion does not change. This particle would still be a causal process, since it *could* enter into interactions, which could make its momentum non-zero. But a shadow wouldn't be a causal process, because it couldn't enter into causal interactions that would make its momentum non-zero. Or, take the following case. Noble gases do not participate in chemical reactions, so they are not involved in a certain type of causal interaction. Is this because they do not, as a matter of fact, exchange the relevant conserved quantities? This is correct, but it cannot be the full answer, for it cannot adequately explain the difference between, on the one hand, a piece of sodium and a piece of chlorine that are never brought in contact and, on the other hand, a piece of argon and a piece of sodium that are never brought in contact. Sodium and chlorine are chemically active. Even if two samples of them were never brought in contact, it is still the case that they would causally interact if they were brought in contact. But to say that argon (and other noble gases) are chemically inert is to say that they wouldn't interact, even if they were brought in contact with a chemically active element. So I think that the full answer should involve appeal to subjunctive (or counterfactual) conditionals: if we were to put sodium and chlorine side by side they would exchange a conserved quantity, but argon and chlorine wouldn't (and couldn't). So it seems that we cannot escape from counterfactuals that easily.

A *second* worry concerns *uninstantiated* causal processes. Causal processes without actual instances do not possess conserved quantities, but *can* possess, and *would* possess them, were they instantiated. Take, for instance, the process of someone's drinking a quart of plutonium. This is certainly a causal process. But it has no instances, since no one has ever drunk that amount of plutonium. Besides, it seems that this causal process cannot have any instances,

since the quart of plutonium is over the plutonium's critical mass. If we are to count it as an unactualized causal process, as I think we should, we need to appeal to counterfactual (and subjunctive) conditionals: were it to be instantiated, it would possess a conserved quantity.

These are still tentative thoughts, and the advocates of the Conserved Quantity theory might well have intelligent replies to offer. So I will close this section with the restatement of the problem. One of the prime advantages of the CQ theory is that it can do away with counterfactuals. But, to say the least, more needs to be said on whether this is really so.¹⁰

4.5 A rough conceptual guide

Before we move on to Part II, we may try to offer a rough conceptual guide to the theories of causation. As noted in the Introduction, there are three ways to divide theories of causation. The first is between generalist and singularist theories. The second is between theories that aim at an intrinsic characterization of causal relations and theories that go for an extrinsic one. The third is between reductive approaches and non-reductive ones.

4.5.1 General vs singular

RVC is generalist, since it makes causation dependent on general patterns (regularities). It does allow a singularist component to causation, since a particular sequence of events can be deemed causal. Yet this singularist component is parasitic on the generalist: a single sequence of events is causal in so far as it is an instance of a regularity. Ducasse's approach (section 2.4), on the other hand, is genuinely singularist, since it takes causation to be fully captured by some local feature of an individual sequence of events c and e , independently of any regularities. Genuine singularists claim that what happened to similar events in the past, or what will happen to them in the future, is totally irrelevant to whether the sequence of events c and e is causal. They *do* allow for a generalist component to causation, but only if it is seen as a generalization over particular causal sequences of events. But there are significant exceptions. As we shall have the opportunity to see in section 6.3.3, Armstrong too

takes causation to be singular, yet he also takes it to be genuinely nomic. Put in a nutshell, his view is that, as a contingent matter of fact, whenever there is singular causation, a causal law is instantiated – where causal laws are not regularities, but relations among universals.

4.5.2 Extrinsic vs intrinsic

Advocates of genuinely singular causation take the view that what determines whether a sequence of events is causal is an *intrinsic* feature of the sequence, whereas the Humean advocates of RVC claim that it is an *extrinsic* feature; namely, that a regularity is thereby instantiated. What is it for a relation to be intrinsic? As with many philosophical notions, there is no simple (and uncontroversial) answer to this question. But we can fix our ideas by saying that a relation is *intrinsic* if the following holds: when two relata stand in this relation, this is entirely a matter of how the two relata are *vis-à-vis* one another, and not at all a matter of their relations to other things. Take the relation *x has more mass than y* as this is applied to the pairs <sun, earth>. It is true that the sun has more mass than the earth and this depends entirely on how the sun and the earth are related to each other and not at all on how these two objects relate to anything else in the universe. Consider now the relation that two objects *x* and *y* have when they belong to the same owner. This is an *extrinsic* relation in the sense that whether it is true that *x* and *y* belong to the same owner will *not* depend on the relation between *x* and *y* but on their relation to a third thing (namely, the owner). With this in mind, we can see that, according to RVC, causation is an extrinsic relation: that event *c* causes event *e* does not just depend on the properties of *c* and *e* and the relations between *c* and *e*, but on their relation to a third thing; namely, a regularity. If then causation is taken to be an *intrinsic* relation, then that *c* causes *e* will have to depend *entirely* on the properties of *c* and *e* and the relations between *c* and *e*.¹¹

It might then be suggested that there is a connection between, on the one hand, generalist and extrinsic characterizations, and, on the other hand, singularist and intrinsic ones. So one might say that Humean theories of causation are generalist and extrinsic, whereas non-Humean are singularist and intrinsic. On a first approximation,

this might be correct. Recall, however, the *human agency* account of Menzies and Price (see section 3.4). There is no reason to think that this is *not* a singularist approach. A human agent could bring about a certain event *c* as a means to bring about another event *e*, even if there was no regularity between event-types *C* and event-types *E*. However, the human agency approach offers an *extrinsic* characterization of causation: whether two events *c* and *e* will be related as cause and effect will depend on a third thing; namely, an agent who manipulates them.¹²

4.5.3 Reductive vs non-reductive

When it comes to the issue of reducibility, things may seem more clear-cut. Recall that the issue is whether causation is ontically autonomous, or whether, instead, it is ontically dependent on non-causal features. There are two ways in which ontic dependence is normally conceived: full reduction and supervenience. Full reductive approaches take the truth conditions of causal statements to be fully specified by reference to non-causal features. One can, however, take a weaker view of ontic dependence, which, strictly speaking, is not reductive; namely, *supervenience*. There are many varieties of supervenience (cf. Kim 1993). But, to fix our ideas, we can think of supervenience as follows: if two worlds are identical *vis-à-vis* their non-causal facts, they are identical with respect to their causal facts too. In other words, the non-causal facts fix all the causal facts. Now, supervenience is a weaker form of ontic dependence than full reduction. If causal facts fully reduce to non-causal ones, then the *prima facie* causal facts are, in essence, *identical* to the non-causal facts. But if causal facts supervene on non-causal ones, then the two sets of facts are *not* identical. Advocates of supervenience, however, take this relation to be strong enough to warrant that the supervenient facts (e.g. the causal ones) are *determined* by the subvenient facts (e.g. the non-causal ones). Lewis, for instance, is a well-known advocate of what he calls Humean supervenience (HS), which is defined as follows: “the whole truth about a world like ours supervenes on the spatiotemporal distribution of local qualities” (1999: 224).

So the idea is this: fix the spatiotemporal distribution of local qualities (which, of course, includes the regularities) and you fix

everything else, including facts about causal relations. Let's then think in terms of supervenience instead of full reduction. The reader should be reminded that there are both generalist *and* singularist supervenience-based approaches. An advocate of RVC takes the subvenient basis to consist of non-causal facts about spatiotemporal relations *and* regularities. A singularist such as Ducasse takes the subvenient basis to consist of spatiotemporal relations between event-tokens and some other non-causal facts, which are *not* regularities (e.g. the fact that an event *c* is the last change before the happening of the effect *e*). More generally, supervenience-based singularists take the subvenient basis of singular causal claims of the form "*c* causes *e*" to involve non-causal properties and relations of *c* and *e*, as well as the local processes that connect them. What a singularist will not allow is that regularities might be part of the subvenient basis. The problem with this stance would be that a subvenient basis without regularities might prove to be very slender to determine all causal facts.

So there can be supervenience-based *singularist* approaches as well as supervenience-based *generalist* ones. Similarly, one might take causation to be an *intrinsic* relation between two events and yet think that this relation supervenes on the non-causal properties and relations of the related events.¹³ It is also open for someone to take causation to be an *extrinsic* relation between two events and yet think that this relation supervenes on the non-causal properties and relations of the related events (i.e. on the regularity that is instantiated by a sequence of events).

All this means that there is conceptual space for someone to *deny* that causation is either reducible to, or supervenient upon, non-causal facts (be they general or singular). This is, in fact, the view taken by Tooley (1984, 1987, 1990). What he calls "causal realism" is the view that "the truth-values of causal statements are not, in general, logically determined by non-causal facts" (1987: 246). So one can take the mark of non-Humeanism to be the denial of HS.

Tooley takes "causal realism" to be a *non-reductive* position, which opposes *both* Humean and singularist reductive (or, supervenience-based) views. Recall that non-reductivism *denies* that the truth conditions of causal statements can be specified in non-causal terms. How are they, then, specified? There are two options available. The first corresponds to what Ehrling (1997: 52) calls "primi-

tivism". This is the view that at least some (basic) causal statements are *not* further analysable in non-causal terms. We have already seen such a view, defended, among others, by Anscombe (section 2.5). Its gist, you might recall, is based on an epistemological argument; namely, that we are directly acquainted with some (singular) causal state-of-affairs, for example, when we observe the cat lapping up the milk. The other option, taken by Tooley (1987, 1990, 1997), is this: although causation *cannot* be reduced to non-causal facts, it can still be *analysed*. Having argued against the view that causal relations can be directly perceived, Tooley (1987: 249) motivates the thought that *causation* can be seen as a "theoretical concept" that can be characterized by means of appropriate theories and postulates.¹⁴ The gist of the non-reductive view is that there is, as it were, causation *all the way down*: there is no way in which causal relations can be specified without reference to irreducibly causal facts.

4.5.4 Humean vs non-Humean

Can we, in light of the three ways of thinking about theories of causation, classify in a neat manner the theories of causation we discussed as Humean and non-Humean? Unfortunately, things are not so straightforward. Take the counterfactual theories. On the face of it, they are singularist, and hence non-Humean. Their basic thought is that causation makes a difference: to say that *c* causes *e* is to say that if *c* hadn't occurred, *e* wouldn't have occurred either. This relation of counterfactual dependence need not have anything to do with general patterns and regularities. Counterfactual dependence is a relation between singular events. Besides, it might well be seen as an *intrinsic* feature of a sequence of events, in virtue of which it is causal. Mackie's version of the theory (section 3.1) is very explicit in its intent to characterize causation as an *intrinsic* relation. However, as we saw in section 3.1.1, Mackie's theory of counterfactuals jeopardizes his attempt to do this. For whether a sequence of events *c* and *e* will be causal will depend on an *extrinsic* feature, namely, the evidence we have for the assertion that if *c* hadn't occurred, then *e* wouldn't have occurred.

Lewis's version (section 3.3) of the counterfactual theory is only *prima facie* singularist, since, in so far as counterfactuals depend on

regularities for their truth, the resulting theory cannot but have an ineliminable generalist component. Besides, Lewis's theory respects HS and, arguably, provides an *extrinsic* way to identify a sequence of events as causal, this being relations of similarity among possible worlds. So Lewis's theory comes out as Humean.¹⁵

Invariance-based approaches, such as Hausman and Woodward's (see section 3.4) are, arguably, quasi-generalist, in the following sense: the very idea of invariance-under-certain-interventions implies repeatability and (at least some) regular behaviour. What remains invariant under interventions is a regularity, even if this is not a universal and exceptionless one. So these theories could count as broadly Humean. They also provide an *extrinsic* way to identify a sequence of events as causal; namely, that the sequence remains invariant under certain interventions. But they fail HS, since the very notion of intervention is irreducibly *causal*, so they can also be seen as non-Humean.

When it comes to mechanistic views of causation, things are even more complicated. They are, typically, concerned with individual causal sequences and, in particular, with what distinguishes them from non-causal ones. So they are interested in *singular causation*. It may well be that the alleged facts about mechanisms are backed up by regularities. But they needn't be, and the mechanisms may well be there to distinguish a causal sequence from a non-causal one, even if there are no regularities that underwrite the singular causal sequence. This is, then, a sense in which mechanistic accounts are non-Humean. Besides, in so far as they deny the Humean claim that there is nothing in the world that forms a tie between causes and effects, that is, in so far as they assert that, apart from spatiotemporal contiguity and regular succession, there is something else in the world (a mechanism or a causal process) that connects cause and effect, they are clearly non-Humean. They also try to locate an *intrinsic* feature of a sequence that makes it causal (e.g. that some structural feature of the process persists, or that a conserved quantity is possessed/transmitted), and this is another sense in which they are non-Humean. But, in so far as the mechanistic views analyse the causal mechanism (or the necessary tie) in non-causal terms, they can be seen as allied to Humeanism. For, along with the Humeans, they espouse HS: they deny that causation is, somehow, an irreducible and fundamental element of the

world. Where they differ from Humean reductive accounts is in their claim that cause and effect are connected not just by a spatiotemporal relation, but also by a local tie.

Perhaps the mark of Humeanism consists in the conjunction of two general theses: first, causation is tied to regularity; secondly, causal facts supervene on non-causal facts. So anyone who denies *either* of these two theses has a non-Humean view of causation. Perhaps only the second thesis (HS) is the true mark of Humeanism. This is not implausible. For, to say the least, there are many philosophers who would endorse HS and yet deny that causation is tied to regularity, simply because they leave room for genuine probabilistic causation. The latter, as noted briefly in the Introduction, is based on the thought that causes make a difference to their effects – whether this difference is such that the effect is rendered certain, probable to happen, or just more probable to happen than not. However, although one might deny that all causation is invariable succession, one cannot deny HS and remain a Humean, for HS is intimately connected with the claim that there is no robust necessity in causation, and in nature in general. And this last claim is, in the end, the bottom line for Humeanism. So we can then say that one is a non-Humean if one takes causation to be an *autonomous* feature of the world, an intrinsic relation among singular events, which cannot possibly be taken to supervene on non-causal facts.

It wouldn't be an exaggeration to say that if causation has a nature at all, it is so complex and multifaceted that none of the theories we have discussed here can fully uncover it on its own. Perhaps, then, we shouldn't crave either a watertight classification of theories of causation into Humean and non-Humean ones, nor a simple and neat theory of what causation is.

II **Laws of nature**

5 The regularity view of laws

5.1 From causation to laws

The Humean RVC ties causation as it is in the world to the presence of regularities in nature: to call a sequence of events *c* and *e* causal is to say that this sequence instantiates a regularity, namely an invariable succession between event-types *C* and *E*. We have already seen, though, that not all regularities establish causal connections. There can be mere correlations of event-types (e.g. the night invariably following the day) that are not causal. So the advocate of RVC, namely, a Humean about causation, should be able to say a bit more about what distinguishes between *good* regularities – ones that can be deemed causal – and *bad* ones – ones that are, as Mill put it, “conjunctions in some sort accidental” (1911: 222). One prominent thought has been that the *good* regularities capture laws of nature, while the *bad* do not. Let us follow customary philosophical terminology and call *accidentally true generalizations* (or *accidents* for short) those generalizations that are true, but do not express laws of nature. In light of this, RVC should be seen as asserting the following two things:

- (a) causation is a species of regularity.
- (b) the species of regularities that causation reduces to are laws of nature.

Let us, now, call the Regularity View of Laws (RVL) the view that:

- (c) laws of nature are regularities.

RVC implies RVL. For (b) above asserts that laws of nature are a species of regularity, hence it implies (c). But RVL does *not* imply RVC. One might accept (c) and yet deny both (a) and (b). Advocates of singular causation, for instance, deny (a), hence they deny that causation reduces to any particular species of regularity; so they also deny (b). But it is open to them to accept that in so far as there are laws of nature, they are regularities – this has been Ducasse’s view – as it is also open to them to argue that there is more to laws of nature than regularities – this is Armstrong’s view. Besides, one might accept RVL but deny that all laws of nature are *causal* laws.

So, the Humean defenders of RVC have to defend RVL. And conversely, although RVL does not imply RVC, a defence of the former will support the latter, for if laws of nature are regularities, then there will be a more solid basis for the view that causation reduces to regularity. Whatever else they are, regularities are recurrent features of the world, and hence, far from being mysterious, causation will be connected to a robust – and objective – feature of the world. All this, of course, provided that the all-important distinction between laws and accidents can be drawn in such a way that no non-Humean commitments follow.

In this chapter, we shall see in some detail how Humeans have attempted to characterize laws of nature and to distinguish them from accidents. But we should not fail to notice that the question “What is a law of nature?” is important in its own right. The language of science is replete with expressions such as *Newton’s laws of motion*, *Maxwell’s laws of the electromagnetic field*, *the laws of ideal gases*, *Mendel’s laws of genetics*, and so on. Besides, there is the widespread view that science aims to discover the laws of nature. And lots of important philosophical issues (e.g. the issue of reduction) turn on the question of whether there are laws in the so-called special sciences as well as on the question of whether there are laws that link physical phenomena with the phenomena studied in the special sciences (e.g. psychological or economic phenomena). Laws are also considered essential to explanation. As we shall see in some detail in Chapter 8, the once dominant deductive–nomological (DN) model of explanation ties explanation of the occurrence of an individual event to its subsumption under a law.

5.2 The Naïve Regularity View of Laws

How are laws different from accidents? The Humean tradition bans objective necessity from nature: there are no necessary connections between events. According to this tradition, there are only regularities, that is, sequences of event-types, which happen in constant conjunction: whenever one occurs, it is invariably followed by the other. When it is said that it is a law that metals expand when heated (under constant pressure), Humeans mean that there is a regularity in nature according to which when a metal gets heated, it expands. There is no necessity in this regularity because (a) it is logically possible that a metal is heated (under constant pressure) and yet it does not expand; and (b) there is nothing in the nature of a metal that makes it the case that, necessarily, it will expand when it is heated. The upshot of this Humean tradition is that both laws of nature and accidents are regularities: they are such that events of certain types regularly succeed each other in space and time. Hence, they cannot be distinguished on the basis that laws establish (or are based on) necessary connections between their constituent properties, while accidents do not. Yet even Humeans are forced to acknowledge that this Naïve Regularity View of Laws faces problems. Even if both laws and accidents are mere regularities, which cannot be differentiated in terms of necessary connections, laws are still sufficiently different from accidents to demand a different treatment. After all, there is a clear intuitive difference between the regularity that all apples in the fruit bowl on the table are ripe and the regularity that all metals expand when heated. Even if all laws are regularities, not all regularities are laws. So *laws are regularities plus something else*. What might this something else be?

In order to fix our ideas, let us formalize the problem by stating that what is at issue is the correct completion of the following formula:

It is a *law* that all *Fs* are *Gs* if and only if . . .

The Naïve Regularity View adds just the following completion:

It is a law that all *Fs* are *Gs* if and only if all *Fs* are *Gs*.

This was part of the early logical empiricist tradition, which took it

that the laws of nature are those regularities that are expressed by *true* universally quantified statements. In fact, the characterization of a lawlike statement was taken to be, on a first approximation at least, purely *syntactic*: a lawlike statement is a universally quantified statement of the form “All *F*s are *G*s”. This view came quickly to grief. Not all statements of this form express genuine laws of nature. Compare, for instance, the statement “All pubs in Britain serve draught beer” with the statement “All metals conduct electricity”. Both of them are true and have the form of a universally quantified sentence. Yet there is a firm intuition that only the second expresses a genuine law of nature. There are a few apparent differences between these two statements, which might help distinguish them, in spite of their common logical form. The first refers to a particular spatial region, namely, Britain, while the second does not. The first has a finite number of instances (since there is a finite number of pubs in Britain), while the second seems to have an unlimited scope. The first has predicates that do not pick out natural kinds, while the second has genuine natural-kind predicates. So, although purely formal criteria cannot adequately distinguish between genuine lawlike statements (i.e. statements that express laws of nature) and true universally quantified statements that do not express laws, there seem to be adequate non-formal ways to ground this distinction.

5.3 Adding sophistication to naïvety

A way to add some sophistication to the Naïve Regularity View is to say that *L* is a statement of a law of nature if and only if:

- *L* is universally quantified; and
- *L* is omnitemporally and omnispatially true; and
- *L* contains only natural-kind predicates, apart from logical connectives and quantifiers (cf. Molnar 1969: 79).

But things are not so easy. One can think of statements that express genuine laws of nature, although they have limited scope and refer to specific spatiotemporal regions. For instance, Kepler’s first law, namely, that all planets in our solar system move in ellipses, is a genuine law; yet it refers to a specific spatial region (our solar system) and has limited scope (there are only nine planets in the solar

system). Besides, one might think of universally quantified statements that make no reference to particular spatial regions, have unlimited scope and genuine natural-kind predicates and yet do not (intuitively) express laws of nature. Reichenbach's (1947: 368) example is very instructive: "All gold cubes are smaller than one cubic mile." This statement has all the features that we have so far demanded of a lawlike statement, and yet it can hardly be said to express a law of nature. More needs to be said about the distinction between laws and merely true universal generalizations.

So the completions of the formula "It is a *law* that all *F*s are *G*s if and only if . . ." suggested so far are not adequate. Humeans must definitely keep the regularity-clause *All F*s are *G*s as part of the full characterization of when it is a *law* that all *F*s are *G*s, for they take laws to be regularities. They must also have legitimate predicates featuring in the generalization. But they must add something else to all this, which is strong enough to distinguish laws from accidents, but also weak enough not to sneak into the concept of a law some suspicious-to-Humeans concept of necessity. Whatever else they take them to be, Humeans must take laws of nature to be *contingent*. In light of this, the problem at hand gets transformed as follows:

It is a law that all *F*s are *G*s if and only if (i) all *F*s are *G*s, and (ii) *X*.

Let's call this hitherto unknown *X* that must be added to a regularity to render it a law the property of *lawlikeness* (cf. Lange 1993: 1). Can Humeans identify the property of lawlikeness?

5.4 The epistemic mark

A prominent attempt to characterize the elusive *X* has been to identify the property of lawlikeness with our different *epistemic attitudes* towards laws and accidents. Lawlikeness is, then, the property of those generalizations that play a certain epistemic role: they are believed to be true, and they are so believed because they are confirmed by their instances and are used in proper inductive reasoning.¹ Braithwaite, who endorsed the epistemic mark of lawlikeness, took the distinction between laws and accidents "to

depend upon knowledge or belief in the general proposition rather than in anything intrinsic to the general proposition itself" (1953: 301).

In particular, he thought that the "honorific title of 'natural law'" should be given to those regularities that are believed to hold on account of diverse evidence in their favour. Braithwaite did take the fact that a lawlike hypothesis occurs in an "established scientific deductive system" as an axiom or a theorem to be relevant to deeming the regularity it expresses to be a law of nature. But he also thought this fact to be just additional *evidence* for taking the regularity to be a law of nature. Take his own example (1953: 302). The hypothesis *All men are mortal* should not count as a law of nature if the only evidence for it comes from its positive instances (i.e. dead men). But if we see it as belonging to an established scientific deductive system, then it also gets support from other hypotheses that entail it (e.g. that all animals are mortal), and it is deemed a law of nature. In sum, Braithwaite thought that to call a generalization a law of nature is to assert: "there are other instances for believing it than evidence of its instances alone" (1953: 302). A. J. Ayer was even more upfront on this point. He (1963: 230) suggested that "the difference between [laws and accidents (what he called "generalisations of fact"))] lies not so much on the side of the facts that make them true or false, as in the attitude of those who put them forward". Elsewhere, he added: "The difference is that when one looks upon a generalisation as a generalisation of law one is willing to extend it to unknown and to imaginary instances in a way that one is not willing to extend any generalisation that one is treating only as a generalisation of fact" (1972: 130).

As Nelson Goodman (1983: 21) famously stated: "rather than a sentence being used for prediction because it is a law, it is called a law because it is used for prediction". So, on the epistemic construal, the expression *It's a law that . . .* should be analysed as follows:

It is a law that all *F*s are *G*s if and only if (i) all *F*s are *G*s, and (ii) that all *F*s are *G*s has a privileged epistemic status in our cognitive inquiry.

Such epistemic construals of the difference between laws and accidents are too subjective and anthropomorphic. Now, this is

exactly the conclusion that Goodman and the rest endorse. But as Ramsey (1928) pointed out, purely epistemic accounts will fail to draw a robust line between laws and accidents. Couched in terms of belief, or in terms of a psychological willingness or unwillingness to extend the generalization to unknown cases, the supposed difference between laws and accidents becomes spurious. For things could have easily been otherwise: we could have been willing to believe or to extend accidental truths instead of true lawlike generalizations.

Another serious problem for this approach concerns the problem of laws that lack any positive instances. Take, for instance, Newton's first law, which says that a body on which no forces are exerted retains its state of motion. This is a law without any instances in the actual world, since there are no bodies in the actual world on which no forces are exerted. But then, it cannot be that this law is a *law* because it is supported (or confirmed) by its positive instances, or because it can be projected to hitherto unexamined ones. There are no such instances to be found. So those who take the epistemic stance face a dilemma: either they have to accept the unpalatable view that Newton's first law is not really a law of nature, or they have to accept the view – equally unpalatable to them – that the mark of lawlikeness is not to do with having instances, and with being projectable to the future on their basis.

In any case, it can be argued that even accidentally true generalizations can be confirmed by their positive instances. Take a theory of confirmation such as Carnap's (1950: 575–5). According to this, if we have a random process of selecting instances from a sample of *F*s, and if all the randomly selected instances have been found to be *G*, then, upon encountering an instance *x* of *F* not included in the sample, the process of random selection will lend support to the claim that *x* will also be an instance of *G*. Now, consider an urn with 10 balls in it, and assume that there is no information about how many of them are red. Suppose that we draw three red balls, and that the mechanism of drawing them is random in the sense that each ball in the urn has an equal chance of being drawn. We can then inductively conclude that all balls in the urn are red. This generalization, if true, is accidental. Still, it is confirmed by its instances, and it can be projected to the cases of the unexamined balls in the urn. It is irrelevant that we could actually withdraw all

the balls from the urn and examine them. Even if we couldn't, the claim that the generalization is confirmed by the withdrawn instances would not be threatened, unless we had reason to believe that the sample was biased.

There is an interesting attempt by Goodman (1983) to connect the thought that only *some* generalizations express laws with the claim that these generalizations contain *natural-kind* predicates. He introduced a new predicate "grue", which is defined as follows: either observed before 2010 and found green, or observed after 2010 and found blue. Clearly, all observed emeralds are green. But they are also *grue*. Why, Goodman wondered, should we take the relevant generalization (or law) to be *All emeralds are green* instead of *All emeralds are grue*? Goodman argued that only the first statement (*All emeralds are green*) is capable of expressing a law of nature because only this is confirmed by the observation of green emeralds. He disqualified the generalization *All emeralds are grue* on the grounds that the predicate *is grue*, unlike the predicate *is green*, does not pick out a natural kind. As he put it, the predicate *is grue* is not "projectable", that is, it cannot be legitimately applied ("projected") to hitherto unexamined emeralds. So whether or not a generalization will count as lawlike will depend on what kinds of predicates are involved in its expression.

No doubt, there cannot be an adequate theory of lawlikeness without a theory of what predicates can be constituents of lawlike statements. Goodman's own theory of *projectability* was couched in terms of the relative "entrenchment" of the predicates that feature in a generalization, where the degree of entrenchment of a predicate is a function of its past uses in projected generalizations. *Green*, for instance, is said to be an entrenched predicate because it has been successfully used in projectable generalizations in the past, but *grue* is not. This idea is interesting, but fails to deliver the goods. For one, it makes the distinction between lawlike and non-lawlike generalizations too subjective. The distinction does not reflect an objective difference between genuine lawlike statements and merely true generalizations, but rather the contingent fact that their constituent predicates have had different uses in the past. For another, Goodman's idea will fail to cover cases mentioned in the previous section. Reichenbach's generalization *All gold cubes are smaller than one cubic mile* seems to have impeccably projectable

predicates, and yet it fails to express a law of nature. As we shall see later on, however, the problem of characterizing what kind of properties are fit for laws will be central to all attempts to characterize laws of nature. Yet, despite its centrality, it has been proved recalcitrant to philosophical analysis.

5.5 Modal force

So far, we have found wanting the attempt to make our epistemic attitudes capture the mark of the lawlike. Ultimately, there seems to be something *objective* in being a law of nature, which is not captured by the epistemic approach. As critics of RVL have repeatedly stressed, laws of nature cannot just be *actual* regularities of the form *All Fs are Gs* precisely because laws issue in “unfulfilled hypothetical propositions” such as “if x were an F , it would be a G ” (cf. Kneale 1949: 75). To say that it is a law that all F s are G s is not just to say that the generalization *All Fs are Gs* has no counter-examples. It’s not, that is, to say that there are no actual exceptions to it. Accidents can well be actually exceptionless. To say that it is a law that all F s are G s is to say that if an object, which is not F , were (had been) F , it would also be (would have been) G . So the critics of RVL point out that it has put the cart before the horse. It is not the fact that a generalization has no exceptions that makes it a law. Rather it is the fact that a generalization *is* a law that deprives it of counter-examples, not just of actual exceptions, but also of *possible* exceptions to it.

Replying to this criticism, sophisticated Humeans have tried to restore some sense in which laws, as opposed to accidents, have *modal force*, by relying on the claim that laws do, while accidents do not, support subjunctive and counterfactual conditionals. We feel, intuitively, that the counterfactual “If this piece of metal had been heated, it would have expanded” is true. But we wouldn’t think true the counterfactual “If this apple had been put in the fruit bowl on the table, it would have been ripe”. This different attitude to the two counterfactuals seems to be based on the strong feeling that only the former is backed up by a law (namely, a real nomological connection between heating a metal and its expansion). The ability of laws to sustain counterfactual conditionals gives some definite meaning to the claim that they have (while accidents lack)

modal force. Take, once again, Reichenbach's example of an accidental generalization: *All gold cubes are smaller than one cubic mile*. This time, let us compare it with the generalization: *All plutonium cubes are smaller than one cubic mile*. Both are similarly unrestricted, and both involve projectable and purely qualitative predicates. Yet we rightly feel that only the latter generalization is a genuine law of nature, because although it is nothing but contingent lack of resources and technical means that do not allow us to create a golden cube larger than one cubic mile, we could not *possibly* create a plutonium cube of this (and of much less) size, even if we had the necessary quantity and means. The construction of this plutonium cube is (physically) impossible because any amount of plutonium over the critical mass would lead to an explosion detrimental to humankind. The plutonium-cube generalization has (while the gold-cube one lacks) modal force. This difference can easily be shown if we take counterfactuals into account. "If this had been a plutonium cube, then it would not have been larger than one cubic mile" is a true counterfactual, while "If this had been a gold cube, then it would not have been larger than one cubic mile" is false. The suggestion then is this: laws are regularities plus *X*, where this *X* is the ability to support counterfactuals. To put it more formally:

It is a law that all *F*s are *G*s if and only if (i) all *F*s are *G*s, and (ii) if an object *x* had been an *F* it would also have been a *G*.

Any attempt to distinguish laws from accidents based on counterfactuals should answer the following question: When exactly is a counterfactual conditional true? Standard truth-functional logic is not suitable for the analysis of the semantics of counterfactuals. A counterfactual conditional (such that "If I had put this sugar-cube in water, it would have dissolved") has a false antecedent (the sugar-cube was never put in water, after all). So, if we were to apply standard truth-functional logic to it, the conditional would be (trivially) true. We need therefore to specify the conditions under which a counterfactual is true and false.² As we saw in some detail in Chapter 3, many philosophers have tried to deal with this difficult issue and considerable progress has been made over the years. But when it comes to the use of counterfactuals in establishing the distinction

between laws and accidents, the news is bad. It turns out that all attempts to establish truth conditions for counterfactuals depend on a prior characterization of what statements express genuine laws of nature. So if we try to distinguish laws from accidents based on the thought that only the former support counterfactual conditionals, we run in a circle. In order to see more clearly how this happens, let us refer briefly to two major theories of the conditions under which a counterfactual can be true.

Goodman (1983) suggests that a counterfactual conditional $p \Box \rightarrow q$ is true iff its antecedent p nomologically implies, given certain other prevailing conditions, the truth of its consequent q . In other words, $p \Box \rightarrow q$ is true iff p , conjoined with some set of facts S and a set of nomological statements L , implies the consequent q : So

$$p \Box \rightarrow q \text{ iff } p \ \& \ S \ \& \ L \rightarrow q.$$

Let us consider a stock example. Take the counterfactual that if this match had been struck, it would have lit. On Goodman's theory, this statement is true because the antecedent (the match is struck), together with other certain facts (e.g. that the match is dry, there is oxygen present, etc.) *and* the laws of nature, imply the consequent (the match lights). The conditions under which a counterfactual is true are inextricably linked with the presence of laws, which determine that, given the antecedent, the consequent must obtain. It's not enough that there is a true universal generalization for the counterfactual to be true. Suppose that it is true that all apples in the fruit bowl are ripe. This is not enough to guarantee the truth of the counterfactual "Had this apple been in the fruit bowl, it would have been ripe". All this means that unless there is already in place a distinction between laws and accidents, Goodman's theory cannot offer an adequate account of the truth conditions of counterfactual conditionals. If, then, we hoped to distinguish laws from accidents by reference to their ability to support counterfactuals, Goodman's account of counterfactuals would not help: it inevitably leads to a tight circle.³

The other important theory of counterfactuals is David Lewis's (1973), which was outlined in section 3.3.1. It suffices for our purposes here to note that according to this theory, the circumstances under which a counterfactual conditional is true involve, at least

partly, laws of nature. So we seem to be moving in the same tight circle that Goodman's theory led us. We first need a distinction between laws and accidents in order to provide Lewis-style truth conditions for counterfactuals. But then it's problematic to appeal to counterfactuals to distinguish between laws and accidents. Criticizing Lewis's approach, L. J. Cohen states this problem neatly:

[Lewis's theory] is capable of elucidating the logic of . . . counterfactuals on the assumption that you are not at all puzzled about what a law of nature is. But if you are puzzled about this, it cannot contribute anything towards resolving your puzzlement. (1980: 219)⁴

5.6 The web of laws

We have been considering some ways to distinguish between genuine laws and accidentally true generalizations and we have seen that the distinction has not been easy to draw at all. In fact, Hempel himself, after noting the "notorious philosophical difficulties" that a proper understanding of counterfactuals faces (1965: 339), admitted that he was unable to offer "a fully satisfactory general characterisation of lawlike statements and thus of laws" (1965: 343). But there seems to be a good way to draw the line. This is known as the Mill–Ramsey–Lewis (MRL) approach but, borrowing Mill's expression, it may be called the *web-of-laws* approach.

Considering how to answer the central problem of "how to ascertain the laws of nature", Mill noted:

According to one mode of expression, the question, What are the laws of nature? may be stated thus: What are the fewest and simplest assumptions, which being granted, the whole existing order of nature would result? Another mode of stating it would be thus: What are the fewest general propositions from which all the uniformities which exist in the universe might be deductively inferred? (1911: 207)

Mill was adamant that he was defending a view of laws as regularities: "for the expression, Laws of Nature, *means* nothing but the uniformities which exist among natural phenomena . . .

when reduced to their simplest expression" (1911: 208). Yet his breakthrough was that the issue of characterizing what the laws of nature are cannot be dealt with by looking at individual regularities and by trying to identify when an individual regularity is a law. Rather, it should be dealt with by looking at how the laws form a "web composed of distinct threads" (*ibid.*). "... the study of nature", Mill suggested, "is the study of laws, not *a* law; of uniformities in the plural number" (*ibid.*).

Criticizing the epistemic view of laws, Frank Ramsey suggested that the difference between laws and accidents cannot be cast in terms of our different epistemic attitude towards them, since, he thought, this difference "would still persist if we knew everything". Then, he (1928: 131) moved on to propose the following: "even if we knew everything, we should still want to systematise our knowledge as a deductive system, and the general axioms in that system would be the fundamental laws of nature". Ramsey's view was revived by Lewis, who (1973: 73) suggested that "a contingent generalisation is a law if and only if it appears as a theorem (or axiom) in each of the deductive systems that achieves a best combination of simplicity and strength".

According to the *web-of-laws* approach: no regularity taken in isolation can be characterized as a law (as opposed to an accident). Lawlikeness is not a property that can be ascribed to a regularity in isolation from other regularities. Laws are those regularities that are members of a coherent system of regularities, in particular, a system that can be represented as a deductive axiomatic system striking a good balance between *simplicity* and *strength*.⁵ Why simplicity? As Ramsey (1928: 131) noted, although the choice of axioms (and hence of the fundamental laws) may sound arbitrary, the requirement of simplicity will certainly constrain this choice. And why strength? Obviously, because the deductive system should be as informative as possible *vis-à-vis* the regularities that hold in the world. But simplicity and strength pull in contrary directions. Hence, a balance should be struck. As Lewis put it:

The virtues of simplicity and strength tend to conflict. Simplicity without strength can be had from pure logic, strength without simplicity from (the deductive closure of) an almanac ... What we value in a deductive system is a properly balanced

combination of simplicity and strength – as much of both as truth and our way of balancing permit. (1973: 73)

A set of true statements can be deductively axiomatized in a number of ways. Of these axiomatizations, some will be simpler than others, while some will be stronger than others. Which system shall we take as the one that expresses the true laws of nature? Lewis's answer to this question is that we should take into account all systematizations that achieve a good combination of simplicity and strength, and that we should take the laws of nature to be expressed by the axioms (and theorems) that are common in all these systems. Notice also that, although we do not know everything, we can idealize a bit. We can conceive of an ideal deductive systematization based on the assumption that we know everything. In any case, as Ramsey suggested,

what we do know we tend to organise as a deductive system and call its axioms laws, and we consider how that system would go if we knew a little more and call the further axioms or deductions there would then be, laws (we think there would be ones of a certain kind but don't know exactly what). We also think how all truth could be organised as a deductive system and call its axioms ultimate laws. (1928: 131)

The useful fiction of an ideal deductive system of the world is not very far from the practice of science as we know it, nor far from what we now take the laws of nature to be.⁶ In any case, if we don't want to speak in terms of the fiction of omniscience, we can claim that there exists a true deductive system of (our knowledge of) the world, irrespective of whether or not we may ever come to know it. To be sure, we know that, if there is one such system, there will also be innumerable such systems. If we are willing to follow Lewis and take the laws of nature to be the axioms and theorems that are common to all these systems (that strike the best balance between simplicity and strength), we can avoid the charge that the web approach to laws requires omniscience.⁷

It follows from the *web-of-laws* approach that accidents are those regularities that do not find a place in the simplest and strongest true deductive system that systematizes our knowledge of the

world. One could, of course, just add all accidental generalizations (what Ramsey called “universals of fact”, as opposed to proper laws, which he called “universals of law”) as extra *axioms* to the best deductive system of the world. But, in doing this, one would make this system far more complicated than it should be. If, for instance, we were to add to the best system Reichenbach’s regularity that all gold cubes are smaller than one cubic mile, we would detract from its simplicity, without gaining in strength. As we have seen, sophisticated versions of RVL take laws to be *regularities plus* X. Now, we seem to have an adequate characterization of the X that should be added to a regularity in order for it to count as a law: laws are regularities in the best systems.⁸ To put it more formally:

It is a law that all *F*s are *G*s if and only if (i) all *F*s are *G*s, and (ii) that all *F*s are *G*s is an axiom or theorem in the best deductive system Φ (or, if there is no unique best deductive system Φ , it is an axiom or theorem in all deductive systems that tie in terms of simplicity and strength).

5.6.1 Strengths, weaknesses and amendments

- The *web-of-laws* approach solves the problem of how to distinguish between laws and accidents.
- It also cuts through the thicket of how best to express this difference in terms of intrinsic features of the statements that express them. According to the *web-of-laws* approach, what makes some regularity a law is *not* an intrinsic feature of this regularity, but rather the fact that it stands in certain relations to other regularities.
- It shows, in a non-circular way, how laws can support counterfactuals, for it identifies laws *independently* of their ability to support counterfactuals.
- It makes clear the difference between regarding a statement as lawlike and being lawlike. In particular, it cuts the Gordian knot that the epistemic approach to laws faced in its attempt to show that laws differ from accidents by virtue of their inductive support and future projectability. Laws might be supported by the evidence, but what makes a regularity a law is not its

relation to the evidence, but its relations to other regularities, as these are expressed in a best deductive system.

- It respects the major empiricist thesis that laws are contingent, for a regularity might be a law in the actual world without being a law in other possible worlds, since in these possible worlds it might not be part of the best system for these worlds.
- It solves the problem of uninstantiated laws. The latter might be taken to be proper laws in so far as their addition to the best system results in the enhancement of the strength of the best system, without detracting from its simplicity. To be sure, empiricists require that laws should be ultimately actual regularities. Their motto for laws could be, as Earman (1984: 210) put it: “in the service of the actual”. So the defender of the *web-of-laws* approach must accept that the uninstantiated laws should play some definite explanatory role within the best system: they must arise in the attempt to account in the strongest and simplest way for actual regularities.

The *web-of-laws* approach constitutes genuine progress in the issue of laws. But there seems to be a major objection we should take care of. It seems that, on this approach, it's not a fact in the world that makes it the case that some regularities are laws while others are accidents. Rather, there is a sense in which it is still our different epistemic attitude towards laws and accidents that marks off the boundaries of the lawlike. For, after all, whether or not some regularities are expressed in the best deductive system (or systems) of the world will depend on how this deductive system is *organized*, and on what is allowed to go into it. Surely, it is an objective fact that a statement is implied or not within a deductive system, a fact that obtains independently of our knowledge of it. But, one may wonder, is this good enough? What statements are so (objectively) implied will indeed depend on the *way* the system is organized. This organization is not necessarily objective. Even if we fix the slippery notion of simplicity, there seems to be no objective way to strike a balance between simplicity and strength. Nor is it guaranteed that there is such a balance. So even if this is not an entirely arbitrary decision, what regularities will end up being laws seems to be based on epistemic criteria and, generally, on our subjective attitude towards regularities, where this attitude is expressed by our

subjective desideratum to organize our knowledge of the world in a deductive system. There is no doubt that the *best system* approach is an improvement over other empiricist attempts to base the distinction between laws and accidents on *overt* epistemic criteria. But many philosophers think that the characterization of laws of nature should be more objective: there must be a worldly feature that makes some regularities laws and others not.⁹

In trying to meet the objection just expressed, we should be careful to distinguish between two issues. The first is whether the *web-of-laws* approach makes laws mind-dependent. The second is whether there is a worldly feature that makes some regularities laws. The reply to the first issue is straightforward. There is nothing in the *web-of-laws* approach that makes laws mind-dependent. As Loewer (1996: 114) has pointedly noted, although, on this approach, the lawfulness of regularities might be mind-dependent, the regularities themselves are not. They are fully objective, and they govern the world irrespective of our knowledge of them, and of our being able to identify them. Besides, there is nothing in this approach that compels its advocates to accept that we cannot be mistaken on what we now take to be the laws of nature. Even in the (fictional) ideal limit of inquiry – where we have availed ourselves of the ideally unified deductive system of the world – we may be mistaken as to what the laws are.

What, then, of the second issue? Is there a worldly feature that makes MRL-regularities laws? In order to answer this question positively, we have to capitalize on some relevant thoughts implied by Ramsey. What, I think, Ramsey was the first to note was that the *web-of-laws* approach *can* offer an *objective* construal of the difference between laws and accidents. In particular, although what regularities are laws will depend on how our knowledge of the world is organized, there is a further fact of the matter that makes laws objectively different from accidents. Discussing the idea of a “best deductive system” Ramsey noted:

what is asserted is simply something about the whole world, namely that the true general propositions are of such forms that they form a system of the required sort with the given proposition in the required place; it is facts that form the system in virtue of internal relations, not people’s beliefs in them in virtue of spatiotemporal ones. (1928: 132)

Ramsey seems to point out (and if he doesn't, I want to point it out) that *it is a fact about the world that some regularities form, objectively, a system*; that is, that *the world has an objective nomological structure*, in which regularities stand in certain relations to each other; relations that can be captured (or expressed) by relations of deductive entailment in an ideal deductive system of our knowledge of the world. Whatever is not, objectively, part of this network of regularities, is an accidental regularity. Conversely, a regularity is nomological if it is part of this network. Ramsey's suggestion is Humean in spirit. Metaphysically, so to speak, laws of nature are regularities: they do not enforce a non-Humean notion of necessary connections. Yet Ramsey's suggestion grounds an objective distinction between laws and accidents in a *worldly* feature: that the world has a certain nomological structure.

5.7 Properties, counterfactuals and accidents

Despite its attractiveness, the MRL view of laws has faced some more criticisms that point to an altogether different conception of what laws are. The first concerns the kind of predicates that are involved in the *best system*. The second concerns the problem of counterfactuals and the third relates to the problem of accidental generalizations. Let us examine them in turn, taking our cues from Armstrong (1983).

5.7.1 Natural properties

Suppose that the best systematization is achieved by Goodman-like predicates such as *grue* (see towards the end of section 5.4). It is perfectly possible that the simplest and strongest deductive systematization may be effected by "unnatural" predicates, that is, predicates that do not pick out natural kinds. Suppose, for instance, that the best system *T* has only two basic axioms; namely, "All *F*s are *G*s" and "All *P*s are *Q*s". Suppose also that predicates *F*, *G*, *P* and *Q* all stand for natural kinds. We could easily create a new system *T'* by re-axiomatizing *T*. The new system *T'* has as the only basic axiom the following: "All *F*s or *P*s are *G*s or *Q*s". We can then introduce two totally artificial predicates *S* and *R* such that *S* = *F* or *P* and *R* = *G* or *Q*. It is difficult to see how an advocate of the *best system*

approach could block the generalization *All Ss are Rs* from being a genuine law of nature, with the original properties (expressed by *F*, *G*, *P* and *Q*) being just artificial divisions of *S* and *R*. But the firm intuition is that it is *not* a genuine law of nature, since the predicates that occur in it do not pick out natural kinds.

Without a theory of what kinds are natural, and hence of what kinds can feature in laws, the advocate of the *best system* approach can only go against this deeply rooted intuition and insist that *All Ss are Rs* is a law. After all, it seems to meet – by hypothesis – the basic condition of the MRL approach: it is an axiom of the *best system* *T'*. It may be objected that *T'* is not as strong as *T*, although it is simpler than *T*. For, in *T*, given that an individual is *F*, we can deduce that it is also *G*. In the new system *T'* however, this extra information is lost since from the fact that an individual is *F* and the lawlike statement “All *Ss* are *R*s” we *cannot* deduce that this individual is also *G*. But, as Armstrong (1983: 68) has aptly observed, this reply by the advocate of the MRL approach would be question-begging. For the point is that had we originally achieved the systematization by means of the artificial predicates, we wouldn't feel that we would lose information by being unable to deduce within the system that “All *F*s are *G*s”. That we feel that way is the outcome of the fact that we know that the new predicates do not pick out natural kinds. The issue at stake, however, is precisely how to distinguish between natural properties and non-natural ones.

The main moral of the first objection is that the requirement of a balanced (in terms of simplicity and strength) deductive system cannot tell us what the laws of nature are, since we also need to know what system captures the nomological relations among *natural kinds*. What we would need would be an objective criterion to distinguish between genuine laws (which express nomological connection among natural kinds) and pseudo-regularities. In pressing this point, Armstrong (1983: 68–9) suggests that a theory of what laws of nature are should deal with the issue of the nature of properties and, in particular, of what properties are natural. Lewis (1983), for sure, accepted this criticism wholeheartedly. For it doesn't seem to damage the *web-of-laws* approach. Its advocate can accept that not all predicates express natural properties and try to explicate the notion of natural property, as Lewis does, in terms of objective similarities and differences between particulars.¹⁰ In any case, it

should be clear that the problem of natural properties is not peculiar to the MRL approach. As we shall see in section 6.3.4, *any* theory of laws, be it Humean or non-Humean, requires natural properties (i.e. properties fit for laws).

5.7.2 Regularities and counterfactuals

The *second* criticism relates to the slippery notion of counterfactual conditionals. We noted above that one of the advantages of the *web-of-laws* approach is that it is able to accommodate the view that laws support counterfactuals, while accidents do not, without the fear of circularity. But do MRL-laws really support counterfactuals? Take, for instance, the law *All planets move in ellipses* and consider what would have happened if the moon were a planet. It is reasonable to claim that it is true that if the moon had been a planet, it would describe an ellipse. In doing so, we have kept fixed the law that all planets move in ellipses. But, as Armstrong (1983: 69) noted, all we are told by the advocates of the MRL approach is that we keep the law fixed in examining the counterfactual because the statement that expresses it is part of the best deductive system. Surely this sounds too subjective. If, for some reason, we were not interested in the best deductive systematization, would the above counterfactual turn out to be false or unwarranted? The intuition behind Armstrong's challenge is that there must be a sense in which the counterfactual "If the moon had been a planet, it would describe an ellipse" is true, irrespective of whether the statement "All planets move in ellipses" is among the denizens of the best deductive system. It is the fact that it is a *law* that all planets move in ellipses that makes the above counterfactual true, and not something about the relation of the statement "All planets move in ellipses" to other statements that express regularities.

Armstrong's moral is along the same lines as the moral of the first criticism: if we take laws to express connections among properties, then we can avoid the supposed interest-relativity of counterfactuals. "If x had been an F , then x would have been a G " is true in so far as the property of being an F and the property of being a G are nomologically connected. This last view has been taken to display an important shortcoming of the *best system* approach – as well as of any version of RVL. Both Armstrong (1983) and Dretske

(1977) point out that, by taking laws to be regularities, RVL fails to *explain* why counterfactual assertions such that “If x were (had been) an F , then x would be (would have been) a G ” are true. Take the MRL-law that *All F s are G s*. This is a *de facto* regularity: it says that all actual objects that are F are also G . But then what makes it the case for an object that is *not- F* that *if* it were an F it would also be G (cf. Armstrong 1983: 69)? Dretske (1977: 266) is even more explicit on this problem. A universal generalization such that *All F s are G s* says “something about the actual F s and G s in *this* world. It says absolutely nothing about those possible worlds in which there are *additional F s* or *different F s*. For this reason it cannot imply a counterfactual” (*ibid.*).

5.7.3 Aren't laws necessary?

The third criticism of the *web-of-laws* approach challenges its claim that it can distinguish adequately between laws and accidents. Critics argue that there may be laws that are not captured by the best system; and conversely, there may be accidents that turn out to be MRL-laws. Imagine, Armstrong (1983: 72) suggests, that there is a set of properties P , Q , R , S , such that it is a law that whenever they are co-instantiated, a new property E emerges. The MRL approach might well fail to capture this law. It may be that the law $(P \& Q \& R \& S) \rightarrow E$ has rare instantiations. Or, it might be that it is quite unconnected to the other laws that hold. So it might well be that adding this lawlike sentence as an extra axiom to the best system increases its complexity without enhancing its strength. The MRL approach would take this to be an accident, although it is a perfectly good law.¹¹

Let's now look at the converse. Van Fraassen (1989: 47) invites us to imagine a possible world in which there are two kinds of shapes and two types of objects. Things in this world are either golden and spherical, or iron and cubic. They are also such that all golden spheres describe circles, while all iron cubes are still. Besides, let us imagine that objects in this world never collide with one another. It would be an MRL-law of this possible world that all (and only) golden objects are spherical. Yet it is hard to see how this is anything but an accident. Slightly different initial conditions might have led some golden spheres to collide with each other, and

hence to have their shapes altered. Although van Fraassen refrains from saying this, the moral here seems to be that there is some sense in which a law is necessary, a sense that is not captured by the *web-of-laws* approach. That all golden objects are spherical would not be a law in this world because it seems perfectly possible that they might not have all been spherical.

Defenders of the MRL approach (e.g. Butterfield 1985: 165) are unimpressed by the latter type of counter-example because they think that our intuitions here are not firm. In some funny worlds (like the one discussed by van Fraassen) it might well be that the genuine laws are what we would be inclined to call accidents in the actual world. But the point remains – and it is certainly an interesting one – that there seems to be some sense in which laws are necessary in a way that accidents are not, and this sense is not captured by the *web-of-laws* approach. We shall turn our attention to this vexed issue in Chapter 6.

Laws as relations among

6 universals

We have already seen that what underwrites the Humean RVL is a disdain for the claim that there are necessary connections in nature. Laws are nothing but contingent regularities plus something else, which distinguishes them from accidents. In this chapter, we shall examine some prominent attempt to show that there is some kind of necessity with which laws of nature hold. It should be noted from the outset that Humeans and many non-Humeans share the intuition that laws of nature are *contingent*. So some non-Humeans try to defend a notion of necessity which is compatible with the view that laws of nature are contingent. They call this *contingent necessitation*. Yet there has been a growing tendency among non-Humean philosophers to argue for a more full-blown account of necessity – *metaphysical necessity* – which makes laws necessary in a much stronger sense. In what follows, we shall have the opportunity to discuss both attempts to show that laws involve some kind of necessity.

6.1 From Kneale's skirmishes to Kripke's liberation war

RVL had been the dominant philosophical view for many decades. Not that there has always been unanimity about it. But up until the 1970s, the few dissenting voices were either not understood, or not taken seriously. William Kneale (1949), for instance, argued that genuine laws of nature differ from merely accidentally true generalizations in a substantial sort of way. Laws, he thought, are not Humean regularities, since laws – as opposed to accidents – hold with necessity. Kneale called laws of nature “principles of objective

necessitation" (1949: 79). He thought that there was a difference between the necessity of some generalizations and the contingency of others, where necessity is "a boundary to possibility" (1949: 78). As examples of how possibility is bounded, he offered cases such as "the incompatibility of redness and greenness" (1949: 80). Not without reason, most philosophers took him to imply that the necessity with which laws are supposed to hold is logical (or conceptual) necessity. The task of dismissing this view was, then, very easy. Lawlike statements – Hume taught us – cannot be logically necessary. After all, the then dominant idea was that to call a statement necessarily true is to say that it's truth is knowable a priori. But surely the truth of lawlike statements cannot be known a priori. A. J. Ayer (1963: 213–14), for instance, took it to be an obvious consequence of the view that laws hold with necessity that "they are purely logical truths" and hence that "they must be discoverable by reason alone". So he summarily dismissed Kneale's view by noting that Hume in his first *Enquiry* established beyond doubt that all laws of nature are known only from experience, and hence they cannot hold with necessity. The moral of Hume's argument, Ayer suggested, was that "there could not be any such relation [of necessary connection] not as a matter of fact but as a matter of logic" (1963: 214).¹

In so far as it is said that there is a concept of necessity that is not logical, then, Ayer thought, it is "mysterious" (1963: 219). This talk of necessity gives no further clues as to how "to detect any laws of nature" (*ibid.*). Unlike Ayer, Popper (1959: 433) did feel the need to capture the intuition that laws of nature hold with (some kind of) natural necessity. He offered the following definition: "A statement may be said to be naturally or physically necessary if, and only if, it is deducible from a statement function which is satisfied in all worlds that differ from our world, if at all, with respect to initial conditions."

This statement, however, is not satisfactory. It amounts to the claim that a statement is naturally necessary if it follows from statements that express laws of nature. Worlds that differ from the actual, if at all, only with regard to initial conditions, are worlds that have the same laws as the actual world. So Popper defines the notion of natural necessity with which natural laws are supposed to hold in terms of natural laws. This is, of course, totally unilluminating. In any case, he quickly retreated to a Humean view of necessity. He said:

I regard, unlike Kneale, “necessary” as a mere word – as a label for distinguishing the *universality of laws* from “accidental” universality. . . . I largely agree with the spirit of Wittgenstein’s paraphrase of Hume: “A necessity for one thing to happen because another has happened does not exist. There is only logical necessity.” (1959: 438)

In a similar fashion, Ayer declared that in so far as a Humean can talk of the *necessity* of a law, this should be taken to mean nothing other than that “there are no exceptions to [the law]” (1963: 220).

So, with the notable exception of Kneale, most philosophers just couldn’t see how laws could be necessary. And having summarily dismissed Kneale’s approach, philosophers such as Ayer (1972: 15–16), Braithwaite (1953: 304) and Popper (1959: 428ff), took it to be the case that laws are contingent.² It was Kripke’s liberating views in the early 1970s that changed the scene radically. By defending the case of necessary statements, which are known a posteriori, Kripke (1972) made it possible to think of the existence of necessity in nature which is weaker than logical necessity, and yet strong enough to warrant the label *necessity*. Besides, Kripke severed the link between a statement’s reporting a necessary truth and its being known a priori. As a result of this, the then dominant view of laws as mere regularities started to be seriously challenged. Armstrong, Dretske and Tooley presented and defended an alternative view of laws that tied laws to the presence of *contingent necessitating relations among universals*, which are knowable a posteriori.

6.2 Against Hume’s Regularity Farm

It’s fair to call the Humean-empiricist RVL reductive in the sense that it reduces laws to regularities. But, as we saw in Chapter 5, laws are not *just* regularities. Something else – the mark of lawlikeness – is also needed to tell *good* regularities (laws) from *bad* regularities (accidents). But the metaphysical status of laws is nothing but regularities. One could paraphrase Orwell’s dictum in *Animal Farm* to capture the Humean thesis: All regularities are equal, but some regularities (the laws) are more equal than others. The Armstrong–Dretske–Tooley (ADT) view of laws can be deemed non-reductive.³ The regularities that hold in the world (or better, a subclass of

them) do not constitute the laws that hold in the world. Rather, and at best, they are the *symptoms* of the instantiation of laws. The so-called accidental regularities differ from the lawlike ones precisely because they are not even symptoms of the instantiation of laws. Take the (toy) law that all ravens are black. On the ADT view, this law does not merely express the regular succession in space and time of two event-types (being a raven and being black). It says something more than that all ravens are black. It expresses a relation between two universals (properties): the property of being a raven (ravenhood) is always co-instantiated with the property of being black (blackness) because there is a relation of *necessitation* between them, which guarantees the co-instantiation. The presence of the regularity that all ravens are black is necessitated by, is explained by, and confirms the relation between the property of ravenhood and the property of blackness.

Where reductivists thought that a law is a special case of a true universal generalization, Dretske noted that there is

an *intrinsic* difference between laws and universal truths. Laws imply universal truths, but universal truths do not imply laws. Laws are (expressed by) singular statements describing the relationships that exist between universal qualities and quantities; they are not universal statements about the particular objects and situations that exemplify these qualities and quantities. Universal truths are not transformed into laws by acquiring some of the extrinsic properties of laws, by being used in explanation or prediction, by being made to support counterfactuals, or by becoming well-established. (1977:253–4)

6.2.1 An intrinsic characterization of lawhood

This lengthy quotation highlights a radical departure from the Humean RVL. The elusive mark of lawlikeness that Humeans wanted to add to (some) regularities to make them laws was an *extrinsic* feature of the regularity. In the epistemic version of the regularity approach (section 5.4), the mark of lawlikeness boiled down to the different role that laws play in induction and confirmation. In the *web-of-laws* approach (section 5.6), this mark was the occurrence of the statement that expresses a regularity in the best deductive system.

The non-reductivists take laws to differ *intrinsically* from regularities in that regularities are not necessarily laws, and conversely, laws do not consist in regularities. The relation of nomic connection, Armstrong suggests, states that one universal necessitates another: “*F*-ness necessitates *G*-ness, or *being F* necessitates *being G*” (1978: 149). If we symbolized $N(F, G)$ the relation of necessitation that holds between *F*-ness and *G*-ness, a law of nature that connects *F*-ness and *G*-ness would simply be $N(F, G)$. Now, Armstrong notes that this relation of necessitation binds *F*-ness and *G*-ness in such a way that when $N(F, G)$ holds, it entails the corresponding (Humean) regularity *All Fs are Gs*. The converse, however, does not hold: the regularity *All Fs are Gs* does not imply that there is a necessitation connection $N(F, G)$. Put in a nutshell, the ADT view is this:

It is a law that all *F*s are *G*s if and only if there is a relation of nomic necessitation $N(F, G)$ between the properties (universals) *F*-ness and *G*-ness such that all *F*s are *G*s.

Although Tooley’s (1977) approach is similar to Armstrong’s and Dretske’s, it deserves a special mention, since it has some interesting implications. Tooley’s main thought is that the best version of RVL, that is the MRL approach, fails to provide adequate truth-makers for law statements. He thinks that there may well be “underived laws”, that is, genuine laws that cannot be derived within an MRL best system. His relevant example is quite instructive.

Imagine a world containing ten different types of fundamental particles. Suppose further the behaviour of particles in interaction depends upon the types of the interacting particles. Considering interactions involving two particles, there are 55 possibilities with respect to the types of the two particles. Suppose that 54 of these interactions have been discovered, one for each case, which are not interrelated in any way. Suppose finally that the world is sufficiently deterministic that, given the way particles of types X and Y are currently distributed, it is impossible for them to ever interact at any time, past, present or future. In such a situation it would be very reasonable to believe that there is some *underived* law dealing with the interaction of particles of types X and Y . . . (1977: 669)

But if, Tooley concludes, there are underived and uninstantiated laws, then it cannot be that their truth-makers are regularities.⁴ Their truth-makers will be “facts about universals” (1977: 672). These facts will consist of “universals’ having properties and standing in relations to other universals” (*ibid.*). So far, so good. But Tooley takes an extra step. He takes it that it is legitimate to generalize from universals’ constituting the truth-makers of uninstantiated laws to universals’ constituting the truth-makers of *all* laws. This move is motivated by the thought that the extension of the claim to *all* laws “provides a uniform account of the truth conditions of laws” (*ibid.*).

The notion of *necessitation* that Armstrong and Tooley suggest is reminiscent of Kneale’s “necessitating principles”. Yet Armstrong is forthright on the claim that this relation of necessitation does *not* amount to logical necessity. It’s not logically necessary that $N(F, G)$ holds. There may be possible worlds in which $N(F, G)$ does not hold. Besides, nomic connections among universals are discoverable only a posteriori. No amount of a priori reasoning could establish that $N(F, G)$ holds. Still, the thought is that in the worlds in which $N(F, G)$ holds, it is the case that *All Fs are Gs* because being an *F* necessitates being a *G*. Armstrong’s master thought is that this relation of necessitation is logically *contingent*. Tooley (1977: 673) agrees with it.

6.2.2 Identification and inference

Humeans will rightly protest that this conception of necessitation is not adequately explained by the non-Humeans. They will note that it is nothing more than a metaphysical rendition of the thought that there must be something that guarantees that a regularity obtains. The problem is really acute – as Armstrong is fully aware. To say that there is a necessitating relation $N(F, G)$ is not yet to explain what this relation is. Nor does it say anything about how the corresponding regularity *All Fs are Gs* obtains. Armstrong insists that $N(F, G)$ *entails* the corresponding (Humean) regularity *All Fs are Gs*; but it is not clear at all how this entailment goes. If the regularity *All Fs are Gs* is contained in $N(F, G)$ as the sentence *P* is contained in the sentence $P \& Q$, then the entailment is obvious. But then there seems to be a mysterious extra *Q* in $N(F, G)$ over the *P* ($=$ *All Fs are Gs*). And we

are in the dark as to what this might be, and how it ensures that the regularity obtains. Note that there is no full agreement among Armstrong, Dretske and Tooley over what this relation of necessitation is. We shall examine Armstrong's views in the next section. For the time being, let us see what Tooley has to say on this issue.

Tooley's view boils down to the claim that the relation of nomic necessitation is a *theoretical construct* (or, better, a theoretical entity). It is introduced in order to explain why a regularity holds, pretty much on the same grounds that theoretical entities are introduced to explain observable phenomena. Now this is a *prima facie* plausible thought, but there seems to be a difference between positing a relation of necessitation and positing, say, electrons. The latter have a certain causal role, which is not exhausted by their role in the production of the phenomena for which they were introduced. Yet the role of the necessitating relation seems to be exhausted in explaining the phenomenon it was introduced to explain; namely, the occurrence of some regularities robust enough to be called laws.

In criticizing both standard (reductive and non-reductive) views of laws, van Fraassen (1989: 38–9) has stated two kinds of issue that any adequate theory of laws should face and solve. He called them “the problem of identification” and “the problem of inference”. Any account of laws must show how to identify laws, and in particular what distinguishes laws from accidents. (It should therefore solve the identification problem.) It must also show how there is a valid inference from the laws there are to the regularities that hold in the world. (It should therefore solve the inference problem.) He then claimed that the reductive account solves the inference problem (since laws *are* regularities), but fails to adequately deal with the identification problem (since, arguably, it does not issue in a robust distinction between laws and accidents). But he also claimed that the non-reductive view of laws fares no better. It solves the problem of identification (since it identifies laws with necessitating relations among universals – which sharply distinguishes laws from accidents). But it fails to solve the inference problem since, even if the so-called necessitating relation among universals is cogent, there is no valid inference from a necessitating relation among universals to the corresponding regularity.⁵ This last objection is particularly pressing for Tooley's account. He (1977: 672) claims that “the fact that universals stand in certain relationships may *logically necessitate*

some corresponding generalisation about particulars, and that when this is the case, the generalisation in question expresses a law". So, on Tooley's view, once it is seen that there is a nomic relation between *F*-ness and *G*-ness, we are entitled to *deduce* from it the corresponding general statement "All *F*s are *G*s". But we are not told *why* we are so entitled. As van Fraassen acutely observes, there is nothing logically compelling in moving from "*F*-ness necessitates *G*-ness" to "All *F*s are *G*s".⁶

6.3 What is the necessitation relation?

Armstrong has made a great effort in order to clarify the notion of the necessitating relation $N(F, G)$ and to solve the inference problem. His main thought is that the relation $N(F, G)$ is *itself* a universal, which is instantiated in the positive instances of laws. Take, for instance, laws such as *All ravens are black*, *All metals expand when heated*, *All planets move in ellipses* and the like. On Armstrong's view they all have forms of the same type: $N(F, G)$, $N(P, Q)$, $N(R, S)$, and so on. They all fall under the type $N(\varphi, \psi)$ where φ and ψ are second-order variables ranging over first-order universals. So the relation of necessitation $N(\varphi, \psi)$ is a *second-order* relation (universal) whose relata are first-order properties (i.e. first-order universals). But why should we take $N(\varphi, \psi)$ to be a universal?

6.3.1 Universals

We have already talked a lot about universals, without explaining what they are. This is a vexed philosophical issue, and one we cannot address properly here.⁷ Suffice it to say that for those philosophers who are realists about universals (that is, properties and relations), universals are really there in the world, as constituents of things (or, better, of states of affairs). Armstrong has been a champion of realism about universals. In his splendid book (1989), he takes universals to be the repeatable and recurring features of nature. When we say, for instance, that these two apples are both red, we should mean, for Armstrong at least, that the very same property (redness) is instantiated by the two particulars (the apples). Redness is a repeatable constituent of things in the sense that the very same redness – *qua* universal – is instantiated in different particulars. One prime reason

for positing universals is the so-called *truth-maker principle*. This asserts that for every contingent truth there must be something in the world that *makes* it true. This principle aims to ensure a certain kind of metaphysical realism: there is something in the world that makes truths true. So there must be a property (e.g. redness) that makes truths about red things true.⁸ As we saw in section 6.2.1, the *truth-maker principle* is a central motivation for positing laws as distinct from regularities, especially for Tooley. Laws are said to be the truth-makers of lawlike statements. Where reductivists take laws to be what true lawlike statements (e.g. the statements of the *best system*) express, non-reductivists take true lawlike statements to be true because there are laws.

Now, it is one thing to accept that properties are (first-order) universals, and quite another thing to accept a whole *hierarchy* of higher-order universals. Yet Armstrong does exactly this. He argues that we need to postulate higher-order properties and relations of first-order properties and relations. One chief reason for this is that the relation of necessitation that he thinks is necessary to account for laws of nature is a higher-order (second-order, in particular) relation between first-order properties. And it is a universal because it satisfies his chief criterion for being a universal: it is a repeatable and recurring feature of nature. As we saw a couple of paragraphs back, the relation of necessitation $N(\phi, \psi)$ is a recurring constituent of all laws. Hence, Armstrong thinks, it is a (second-order) universal. Admitting that $N(\phi, \psi)$ is a universal, argues Armstrong, can lead us to see how a specific necessitating relation $N(F, G)$ is such that it *guarantees* that the corresponding regularity *All Fs are Gs* obtains. So he thinks he can solve the sticky *inference problem* (see section 6.2.2).

6.3.2 Necessitation and causation

Armstrong's solution is complicated, but it can be broken down into four steps.

1. We can have direct (i.e. non-inferential) causal knowledge when singular causings occur, for example, when one feels the pressure of one's own body. This knowledge comes from perceiving causings. At least some of these perceptions are "as epistemically primitive as any other perception" (1993a: 421).

2. Because such singular causings fall, typically, under patterns, where the same type of cause F produces the same type of effect G , we infer the existence of universals (F -ness and G -ness) to explain the sameness in the instances of the pattern; namely, that all F s are G s. That is, we infer that there are universals (properties) such that they are co-instantiated in the pattern at hand.
3. We then move on to explain the very co-instantiation of the universals in the pattern by positing the existence of a necessitating relation $N(F, G)$ among the universals F -ness and G -ness, which guarantees their co-instantiation. We posit, that is, that there is a higher-order relation $N(F, G)$ between the first-order universals F -ness and G -ness.
4. This relation $N(F, G)$ is a *causal* relation. In fact, Armstrong (1993a, 422) says of $N(F, G)$ that “it is the very same relation that is actually experienced in the experience of singular causal relations, now hypothesised to relate types not tokens”.

Given these steps, it is not difficult to see why Armstrong thinks that the *inference problem* is solved, that is, that the regularity *All F s are G s* is entailed by $N(F, G)$: if the relation $N(F, G)$ is a *causal* relation between F -ness and G -ness, then each and every token of type F causes a token of type G . “The inference”, as Armstrong states, is “analytic or conceptual”. Put in a nutshell, the idea is this:

$N(F, G)$ implies “All F s are G s” because

- a) $N(F, G)$ implies *for all x , $N(Fx, Gx)$* , where $N(Fx, Gx)$ means that x ’s being F necessitates x ’s being G ; and
- b) *For all x , $N(Fx, Gx)$* implies “All F s are G s”.

Although Armstrong anticipated this solution earlier (1983: 88), there he stated that, at the end of the day, “the relation of nomic necessitation, N , will have to be accepted as primitive”. In Armstrong (1993a), he implies that the relation of nomic necessitation is (a) hypothesized to hold in an attempt to offer the best explanation of regular patterns in nature; and (b) that, at the end of the day, it is a *causal* relation. With regard to (a), he came quite close to Tooley’s view that was outlined towards the middle of section 6.2.2. So the problem that was raised there applies no less to

Armstrong's view. With respect to (b), Armstrong made a novel point, but it is controversial and problematic. Even if we were to accept that universals could cause other universals, still the first step of Armstrong's foregoing argument, namely, that we have direct perceptions of causings, will be at best question-begging against the Humeans. For, as we saw in section 2.5, they would deny that there are such things as singular causings, let alone that we can perceive them directly – *qua* causings.⁹ So Armstrong has to take a radical non-Humean line as the first step of his argument for necessitation, but we would need an independent reason in order to take seriously the view that we can directly perceive causings. And, as we saw in section 2.5, this reason is not forthcoming.¹⁰

6.3.3 Nomic singular causation

In recent writings, Armstrong makes the connection between causation and lawhood even tighter. He notes (e.g. 1997: 507) that the relation of nomic necessitation is, in fact, the causal relation. He remarks that the “*N*” in $N(F, G)$ should give its place to a “*C*” for *cause*. Given his view that we can have direct experience of singular causation “in pressure on the body and the operation of the will”, he (1993b: 173) invites us to view “the nomic relation between universals . . . as being the same as (or analogous to) token [i.e. singular] causation of this experienced sort”.

This direct identification of necessitation with causation is not without problems. Already in the fourth step of his argument, which was presented in the previous section, Armstrong assumes that the causal relation – C_1 – that exists between two singular events (e.g. the striking of this match and its lighting) is the same as the causal relation – C_2 – that exists between two universals (e.g. the universal *striking a match* and the universal *lighting of a match*). But, as van Fraassen (1993: 436) has asked, what is to guarantee that $C_1 = C_2$? To just posit that $C_1 = C_2$, as Armstrong in effect does, does not offer an adequate explanation of the supposed identity. In any case, suppose that $C_1 = C_2$. It is still not clear why the inference problem is solved by anything other than by fiat.

What is also worth noting is that, in the end, Armstrong tries to bring together singular causation with the presence of laws. This is not accidental, of course. This move is required by his thought that

a causal relation between two event-tokens and a nomological relation among universals is the *very same relation*. So he takes it that singular causation should be identified with the instantiation of a law, but this identification is made on empirical, a posteriori grounds (cf. Armstrong 1997: 507, Heathcote & Armstrong 1991). Being non-Humean about causation, that is, rejecting RVC, Armstrong takes it to be the case that there is singular causation. In Armstrong (1983) he argued that causation is not essentially nomic. But in his recent writings, and especially in the article he coauthored with Heathcote (cf. Heathcote & Armstrong 1991: 66), he suggests that this is “a profoundly unsatisfactory position”. Does that then mean that Armstrong has now yielded to the Humean view that all causation is nomological? The answer is partly yes and partly no. It is positive in the extent to which Armstrong thinks that where there is causation there is instantiation of laws. But it is negative in the extent to which he thinks that these laws are non-Humean laws: they embody relations among universals, and they are not mere regularities.

A distinctive feature of Armstrong's view of laws is that he takes all universals to be instantiated and to “have no existence, except in [their] instantiations” (1997: 506).¹¹ So there is no property of, say, redness, if there are no red things; the relevant universal (redness) exists wholly in red things. Armstrong entertains this view because he does not want to place universals outside the spacetime realm. So it is his *naturalism*, which, although independent of his realism about universals, dictates that although there are universals, they are not the traditional abstract entities, which exist outside space and time (cf. 1983: 82). But then he thinks that it is easy for him to retain the strong non-Humean intuition that there are singular causes, and at the same time accept the view that “singular causation is nothing but the instantiation of a law” (1997: 506–7). The combination is effected thus: “Since, like any universal, it is complete in its instance, the law is complete in its instance. At the same time, because it is a universal, it is potentially general: It may be found in other instances, and so can function as a general law” (*ibid.*).

Here again, the Humeans may reply that (a) one of their main points – that there is no non-nomic causation – has been granted; and (b) the addition that Armstrong makes – that nomic causation is a universal – plays no *extra* role over the Humean admission that

nomie causation is, in the end, regular succession. Since, however, these are new and intricate developments, I shall not pursue them further. I will only note a problem that needs further examination. Armstrong's recent views seem to tie too close a knot around the concepts of law and causation. He seems to be committed to the view that all laws (being relations among universals) are causal laws. This is, surely, an open issue. It seems that there are laws of which one can argue that they are not causal (e.g. laws of coexistence, or conservation laws). The MRL approach seems to be able to accommodate non-causal laws in its scheme, since non-causal laws may well be axioms or theorems of the best deductive system. But Armstrong is in need of a criterion to distinguish which relations among universals are causal and which are not, if he is to accommodate non-causal laws in his own scheme.¹²

6.3.4 Natural properties revisited

In any case, one can question whether Armstrong's appeal to a necessitating relation among universals can issue in the distinction between laws and accidents. Why is it the case that *All apples in my fruit bowl are red* is not a law, whereas *All metals expand when heated* is? In other words, why isn't there a necessitating relation that issues in the first regularity, but there is one that issues in the second? Here, the quick answer would be that *being an apple in my fruit bowl* is not a universal, since it refers to a particular (my fruit bowl). Yet *being an apple in my fruit bowl* is repeatable, and hence it seems to have what it takes to be a universal. Indeed, Armstrong (1983: 100–1) considers a similar objection and, in order to address it, he introduces the notion of "quasi-universal" to accommodate the claim that some seemingly accidental generalizations can, nonetheless, be laws. But isn't this move too strong? If we were to follow it, then all kinds of generalizations could be seen as instances of necessitating relations among quasi-universals. Of course, Armstrong does not want this. It is not clear, however, how he can avoid it. He would invite us to counsel our intuitions to see whether a generalization can or cannot be deemed a law, in his sense of law. He could invite us, for instance, to see that although being made of metal *causes* a rod to expand when heated, *being an apple in my fruit bowl* does not cause it to be red. One can wonder how trustworthy these intuitions

may be. If, at the end of the day, we force it to be the case that the necessitating relation that makes something a law is exemplified only in the cases in which our intuitions say there are genuine causal relations, we may end up with nothing but an *ad hoc* manoeuvre as a way to distinguish between laws and accidents.

The broader problem here is that Armstrong, no less than the advocates of MRL and other Humeans, needs a theory of what kinds are natural (i.e. of what universals express natural properties) and of how genuine nomic relations pertain only to such kinds. Armstrong (1978) has tried to offer such theory. Commentators (e.g. Carroll 1990: 200–1) argue that it is fraught with problems. Be that as it may, three points need to be stressed. First, the theory of what universals are natural should be based on objective – and not epistemic – considerations, if Armstrong’s view is *not* to fall prey to the criticisms that he himself has offered against the Humean approach to laws. Secondly, it should be independent of what universals are taken to feature in examples of genuine laws, if it is to offer a non-circular account of what laws of nature are. And thirdly, the very need of the non-reductive view to appeal to natural properties suggests that the similar need of the reductive approach should not, on its own, be taken to count against the view that laws are regularities.¹³

Drawing this long discussion to a conclusion, we may note that many philosophers think that, despite its undeniable ingenuity, Armstrong’s appeal to a necessitating relation does not really make a case for the supposed difference between laws and accidents in terms of necessary connections. It seems unclear that the necessitating relation is anything other than an *ad hoc* manoeuvre. It fails to guarantee that there is something more to laws than constant conjunctions among event-types.¹⁴

6.4 A trip to the metaphysical heaven

The advocates of the ADT approach to laws unite with the Humeans in one central thesis: that laws of nature are *contingent*, that is, that there are possible worlds in which they do not hold. Recently, however, there has been a growing tendency among philosophers to think of laws of nature as *metaphysically* necessary: they hold in all metaphysically possible worlds. One motivation for

this reaction is the thought that, by making the necessitating relation contingent, the standard rendition of the ADT view fails to capture a robust sense of necessity. And where reductivists would take comfort from that in their attempt to show that no robust sense of necessity has been defended, these philosophers argue that the failures of the ADT approach is reason enough to look for this more robust sense of necessity – *metaphysical necessity* – in virtue of which laws hold. According, then, to Crawford Elder (1994: 649) laws hold with “full strength necessity”. Or, as Chris Swoyer (1982: 222) has put it, “laws of nature are metaphysically necessary relations between properties”.

Now, metaphysical necessity is distinct from logical necessity. But in exactly what sense? Logical truths are necessary in the sense that they hold in *all* possible worlds. Can we say the same about the laws of nature? We have already seen (section 6.1) that this thought had been taken to be absurd for two reasons: first, it was thought that laws are contingent, so that they cannot be logically necessary; secondly, it was thought that if laws were logically necessary, then they would be knowable a priori, which is absurd. But after the Kripkean revolution, it was taken to be the case that there can be truths that hold in *all* possible worlds but are knowable a posteriori. These might be taken to be the metaphysically necessary truths. For instance, as Kripke has famously argued, the identity *water* = H_2O , in so far as it is true, is metaphysically necessary and yet discoverable a posteriori. To be more precise, the former identity is said to hold in all possible worlds in which water *exists*. So the statement “water = H_2O ” is not logically (or conceptually) true. It should then be stressed that metaphysically necessary truths are said to be grounded in the natures of things (and not in our concepts about them): they hold in all possible worlds in which the things (entities, substances) they are about exist.¹⁵

If laws of nature are metaphysically necessary, then they too should be grounded in the nature of things involved in them. That is to say, given that some things have the nature (properties) they do, they *must* obey the laws they do. Swoyer, who was one of the first to press this view, thinks that the relation of necessitation can be analysed away. If we take account of the intrinsic nature of properties, he suggests, we can conclude that two (or more) properties “give rise to the law without the need for any additional relation”

(1982: 218). How, then, does the law arise? We don't get much of a clear picture here. Indeed, the issue of how there can be a necessitating relation among properties is just pushed one step back. Where the contingentist necessitarians thought that there are properties and (contingent) necessitating relations among them that give rise to laws, Swoyer says that (some) *F*-ness cannot be instantiated without *G*-ness also being instantiated because "the very natures of [*F*-ness] and [*G*-ness] make it impossible for there to be a world in which it is false [that All *F*s are *G*s]" (1982: 216–17). I submit to the reader that I don't find this claim particularly enlightening.¹⁶

Having said this, however, there has recently been an important attempt to ground the claim that laws of nature are metaphysically necessary; namely, *dispositional essentialism*. This view has been vigorously defended by Brian Ellis and Caroline Lierse (cf. Ellis 1999, 2000, 2001, Ellis & Lierse 1994). Its thrust is the idea that natural kinds have dispositional essences, that is, causal powers, capacities or propensities, which they possess *essentially* and in virtue of which they are disposed to behave in certain ways. So, for instance, water has essentially the causal power to dissolve salt and it is in virtue of *this* power that it does dissolve salt in the actual world and that it is a *necessary truth* that water dissolves salt. As Ellis (2000: 344) puts it: "The causal laws are not contingent universal generalisations about how things actually behave, but necessary truths about how they are intrinsically disposed to behave."

On this view, laws of nature are ontologically dependent on the intrinsic natures (essences) of natural kinds: given that the natural kinds are essentially what they are, and given that they are thus intrinsically disposed to behave in certain ways, the causal laws they give rise to are fixed. *Dispositional essentialism* constitutes an important break with the Humean metaphysics. To say the least, it challenges the basic Humean assumption that laws of nature supervene on non-modal facts. It also reverses the main Humean order of dependence: where Humeans think that it is the properties of an entity *and* the laws it obeys that determine the natural kinds to which this entity belongs, dispositional essentialism argues that it is the essential properties of a natural kind (some of which are inherently dispositional) that determine what objects fall under it and what laws they must obey (cf. Ellis 2000: 340–41). But many philosophers will find dispositional essentialism unappealing, not

least because it seems to fail to explain how (and in virtue of what) there is this supposed fundamental distinction between essential and non-essential properties.¹⁷

6.5 Humean nomic supervenience

Earman (1984: 195) has suggested that the “empiricist loyalty test on laws” is a supervenience thesis: there can be no difference in laws, unless there is a difference in non-nomic (Humean) facts. So imagine two worlds W_1 and W_2 that agree on all non-nomic facts. Then, this supervenience thesis implies that W_1 and W_2 will also agree on laws. Let’s call this thesis *Humean nomic supervenience* (HNS). On a good approximation, we can take these non-nomic facts to include regularities, as well as psychological facts, or facts concerning the simplicity and strength of a deductive system. It is clear that the Humean accounts of laws examined in Chapter 5 satisfy HNS. It’s equally clear that the ADT account of laws does not satisfy HNS.

In a series of papers and a book (1987, 1990, 1994), Carroll has tried to show that HNS is false. His argument is based on modifications of Tooley’s example, which was presented in section 6.2.1, and which tries to make a case for the existence of underived and uninstantiated laws. Carroll invites us to consider two possible worlds W_1 and W_2 , which agree on all non-nomic facts. In both W_1 and W_2 there have never been, and will never be, fundamental particles of type X , nor fields of type Y . However, Carroll says, there is an (uninstantiated) law L_1 in W_1 that governs the interaction between field Y and particle X . According to L_1 , all X -particles subject to Y -fields have spin *up*. W_2 is exactly like W_1 except that in W_2 law L_2 holds instead of L_1 . According to L_2 , all X -particles subject to Y -fields have spin *down*. So, Carroll argues, although the two worlds W_1 and W_2 differ in their laws, they do *not* differ in any non-nomic facts. So HNS must fail. To strengthen his case, Carroll also invites us to suppose that neither in W_1 nor in W_2 are there any cognizers, whose psychological or epistemic attitude might have helped them to adduce non-nomic facts to characterize the laws in the two worlds. His conclusion is that the very possibility of two worlds agreeing in all non-nomic facts but disagreeing in laws (a possibility that is grounded in his thought-experiment) shows that laws do *not* supervene on non-nomic facts.

Before we examine Carroll's argument, let us note that it is *not* his conclusion that there are facts about universals that determine laws. Carroll is no friend of the ADT view. In fact, he argues (Carroll 1987) that the latter should fail as an account of laws no less than Humean accounts do. Curiously enough, Carroll takes the ADT view to be a reductive account of laws, where the reductive basis consists of facts about universals. I think this is wrong. For the facts about universals that determine laws in ADT include nomic facts, that is, facts about the nomic necessitating relation that makes the property of *F*-ness bring about the property of *G*-ness. Be that as it may, what is at stake here? Having (presumably) shown that HNS fails, and having shown that the alleged reductive account of ADT also fails, Carroll (1987: 266) suggests that the only position that is left open is a non-reductive view of laws according to which laws are "primitive and irreducible". To take the latter view is, according to Carroll (1987: 267), to deny that law statements have truth-makers, and hence to refrain from any further analysis of the supposed facts (either nomic or non-nomic) that are the truth-makers of laws. At the same time, Carroll thinks that not only are there laws, but those philosophers who think there are not are mistaken.

So is Carroll's thought-experiment against HNS conclusive? Hardly. There is, first of all, the issue of how seriously we should take the thought that Carroll's example presents *real* possibilities. What is at stake here, as Loewer (1996: 116) has put it, is how reliable our intuitions are. The example relies on us finding it intuitively possible that there are two worlds W_1 and W_2 that are alike in all Humean facts but have conflicting laws. Since there is, by hypothesis, no way in which this difference in laws could show in facts of these worlds, one can wonder whether the intuition that there is such a difference can be trusted. A similar concern is voiced by Earman (1984: 210–11). He also points out that the thought-experiment at hand is consistent with the MRL approach to laws. As we saw in section 5.6.1, the MRL approach does allow for uninstantiated laws, provided that their inclusion in the best system enhances its strength, without detracting from its simplicity. In the thought-experiment we are discussing, there are two mutually incompatible uninstantiated laws that hold in two different, but Humean-indistinguishable, worlds. However, the Humean could

argue, had these laws been instantiated they would have given rise to different regularities. That is, there would have been an actual (Humean) difference in the two worlds. Hence, which of the two laws holds would be determined by non-nomic facts in worlds W_1 and W_2 .

It might be heartening to the defender of HNS that Armstrong (1983: 123–6) himself doubts that Carroll's (and Tooley's) counter-example commits us to the existence of genuine uninstantiated laws. For him, uninstantiated laws are "concealed counterfactuals" of the form: "if, contrary to fact, certain sort of things existed, then these things would obey a certain law". The truth or falsity of such counterfactuals "depends wholly upon actual, that is instantiated, laws" (1983: 126). So, even with a little help from Armstrong, HNS is not threatened by Carroll's counter-examples.¹⁸

Even if HNS failed, why, one may wonder, should we accept Carroll's positive view? The way he motivates it suggests that we gain nothing by trying to explain what makes some law statements true. Carroll says: "nothing explanatory is gained by positing relations among universals as truth-makers" (1987: 266). On the contrary, he claims, by taking lawhood as primitive (and certainly non-reducible), we do not inflate our ontology. We don't go for a trip to the "metaphysical heaven" (*ibid.*). Armstrong once said that we must admit the existence of nomic necessity "in the spirit of natural piety" (1983: 92). Yet he went on to say a lot more by way of analysis. Carroll's treatment of nomic necessity as primitive just calls us to be pious, without even trying to explain to us why we should. Why such devotion, as Loewer (1996: 119) naturally wondered? If we had a definite gain by this pious act, then we might consider undertaking it. But we haven't. We don't even have an account of how, by taking laws to implicate an unanalysed notion of nomic necessity, they explain anything. At least, with the MRL approach we can understand how laws *explain*: they do so by unifying. In Chapter 10, we shall see in detail how in fact the MRL approach to laws can be seen as the next of kin to those theories that view explanation as unification. But before we move on to *explanation*, let us examine some alternative approaches to laws of nature.

Alternative approaches

7 to laws

In recent years, there have been a number of alternative approaches to the characterization of laws of nature. They are quite different from each other, but they all unite in claiming that we are not forced to choose between the Humean RVL and the ADT view that laws are necessitating relations among universals. In this chapter, we shall examine the most prominent of them.

7.1 Methods and inference-tickets

After reviewing some standard Humean attempts to distinguish between laws and accidents, L. J. Cohen (1980) suggests that they wrongly pay too much attention to the ill-motivated task of defining the conditions under which a statement expresses a law, while they disregard the wider epistemological context in which law statements occur. They look to semantics for an elucidation of the distinction between laws and accidents, but, Cohen argues, the real elucidation comes from looking at the *role* that laws and accidents play within “cognitive inquiry” (1980: 222). His basic thought is that what distinguishes between laws and accidents has nothing to do with the content of the statements that express them, but, instead, with the *methods* by which certain statements are arrived at and are justified. Cohen contends that statements that are rightly deemed to express laws of nature are the products of the method of *eliminative induction*, whereas accidentally true generalizations are the products of implementing *enumerative induction*. This might suggest that there is no sharp difference between laws and

accidents. But Cohen is happy with this. As he (1980: 223) puts it: "A real law of nature is, as it were, the crock of gold at the end of the inductivist rainbow."

Now, as Cohen construes it, enumerative induction is the method by which a generalization is established after examining its instances. The enumeration of the instances need not be exhaustive. Nor should the generalization have only a finite number of instances. Rather, Cohen's point is that the generalization that gets produced by enumerative induction receives support *only* from its favourable actual instances. In contrast to this, eliminative induction is the method by which "we test the capacity of a low-grade generalisation to resist falsification by varying the experimental circumstances in which it is tested, and we test a high-grade scientific theory by its capacity to explain a variety of accepted lower-grade uniformities and predict some new ones" (1980: 223). So when a generalization is established by means of eliminative induction, it gets supported not just by its positive instances, but also by collateral and variable evidence. Or so the idea is. When we have such a generalization, Cohen notes, "we have as good a reason as we ever have for calling it a law" (*ibid.*).

A natural worry here is that Cohen has not really offered an elucidation of the distinction between laws and accidents, which is, in principle, different from that offered by some Humeans such as Braithwaite and Ayer. As we saw in section 5.4, they too have argued that the distinction between laws and accidents is best understood as expressing how the relevant generalization gets supported by the evidence. To be sure, Cohen's insight is that there is a "close correlation between the nomological-accidental distinction . . . and the eliminative-enumerative distinction . . ." (1980: 224). But even so, is this correlation robust enough to warrant the view that there is a royal (if defeasible) road to lawhood? It might well be that the more variedly corroborated a generalization is, the more confident we should be that its truth, if it is true, won't be accidental. But here again, intuitions are not firm enough to clinch the issue. In any case, it is debatable that there is such a deep difference between the two methods as the one suggested by Cohen. Enumerative induction, where we just enumerate instances and form a generalization, will either be an untrustworthy (and, in the end, a naïve) method, or else it will be just a species of eliminative

induction. For in good applications of enumerative induction, one should, for instance, eliminate alternative hypotheses, for example, that the sample is biased, before one draws the generalization from the observed instances.¹

Cohen advanced his own position (at least partly) as a reaction to the once popular inference-ticket view of laws. The thrust of this view is that law statements should not be seen as expressing propositions, and hence as being amenable to claims of truth and falsity. Rather, they must be seen as disguised *rules of inference*. We cannot validly move from the singular claim that “*a* is *F*” to the singular claim (perhaps, prediction) that “*a* is *G*”, unless we use the sentence “All *F*s are *G*s”. On the inference-ticket view, the function of law statements is exactly this: they entitle us to make inferences such as the above.

This view has a venerable history. Schlick and Ramsey were among its advocates. Schlick’s (1931) endorsement was motivated by the thought that nomological statements are, strictly speaking, meaningless, because they are unverifiable. He thought, however, that although meaningless, nomological statements can provide the major premise in arguments whose minor premise and conclusion are verifiable.² In a similar fashion, Ramsey (1929: 137) suggested that “causal laws form the system with which the speaker meets the future”. They “are not judgements but rules for judging ‘If I meet a ϕ , I shall regard it as a ψ ’ . . .”. The inference-ticket view came to grief for three reasons. First, the empiricists’ demand of strict verifiability as a criterion of meaningfulness was deeply problematic. Secondly, it is questionable that the statement “All *F*s are *G*s” can serve as a premise in a valid deductive argument without having a truth value. But if it is endowed with a truth value, then it expresses a proposition, and hence it has content. This content can be naturally said to be the law it expresses, if it is taken at face value.³ Thirdly, the inference-ticket view patently fails to account for the difference that there is between laws and accidents in supporting counterfactuals. Take the generalizations *All metals expand when heated* and *All coins in my pocket are dimes*, where the first is lawlike, while the second is accidental. Both of them can be premises of valid arguments, whose other premise is “*a* is *F*” (i.e. *a* is a coin in my pocket/ *a* is metal) and whose conclusion is “*a* is *G*” (i.e. *a* is a dime/ *a* expands when heated). If they are both seen as

inference-tickets, then it is a mystery why we would be entitled to say that *if a were a metal and was heated it would expand*, whereas we are *not* entitled to say that *if a were a coin in my pocket, it would be a dime*.

7.2 Intervention and invariance

In a seminal piece, Woodward (1992) suggested that the elusive notion of lawfulness should be linked to the notion of invariance. Where Humeans tried to avoid any concept of necessity, and non-Humeans tried to reply to them by defending a metaphysical notion of necessary connection, Woodward thinks that there is space for the defence of a non-Humean view of laws, which (a) avoids unnecessary metaphysical commitments, and (b) stays close to scientific practice. Think, Woodward says, of how the notion of lawfulness is used in science. It is “closely connected with the notions of stability and invariance” (1992: 202). Take for instance the law of the pendulum: $2\pi\sqrt{l/g}$, where T is the period of oscillation and l the length of the rod. To say that this equation expresses a law is to say that the relationship between the period T and the length l “will remain stable or invariant under some fairly wide range of changes or interventions” (*ibid.*). So the relationship will remain the same if we move the pendulum to a different place, or if we change the length of the rod, or if we change its material, and so on. If this were an accidental generalization, Woodward invites us to see that we would expect its invariance to change to be either absent, or very limited. In fact, we would expect that although the generalization truly describes the relationship between the actual values of the magnitudes T and l , it wouldn’t hold for different, or merely possible, values.

In subsequent work, Woodward (1997, 2000) has tried to elaborate on the view just sketched and, in particular, to analyse further the central notions of invariance and intervention. Suppose that we want to know whether two variables (or magnitudes) X and Y are related causally or nomologically. On a first approximation, this is achieved by finding out whether “the intervention on X produces a corresponding change in Y ” (1997: S30). Woodward’s characterization of an intervention is too technical to be properly explained here. But the gist is this. A change of the value of X counts as an intervention I if it has the following characteristics:

- (a) the change of the value of X is entirely due to the intervention I ;
- (b) the intervention changes the value of Y , if at all, only through changing the value of X .

The *first* characteristic makes sure that the change of X does not have other causes (other than the intervention I , that is), while the *second* makes sure that the change of Y does not have causes other than the change of X (and its possible effects).⁴ These characteristics are meant to ensure that Y -changes are exclusively due to X -changes, which, in turn, are exclusively due to the intervention I . For example, if we wanted to find out whether the length of a pendulum (X) is nomically related to its period (Y), we would have to make an intervention I whose only impact is the change of the length X . If the period Y also changed as a result of this intervention, and if Y 's change was only due to the change of X , then we could say that there is a nomic relation between the length of a pendulum and its period. As Woodward notes, and as we have already seen in section 3.4, there is a close link between intervention and manipulation. Yet his account makes no special reference to human beings and their (manipulative) activities. In so far as a process has the right characteristics, it counts as an intervention. So interventions can occur *naturally*, even if they can be highlighted by reference to "an idealised experimental manipulation" (2000: 199).

Having analysed the notion of intervention, Woodward proceeds to link it with the notion of invariance. A certain relation (or a generalization) is invariant, Woodward says, "if it would continue to hold – would remain stable or unchanged – as various other conditions change" (2000: 205). What really matters for the characterization of invariance is that the generalization remains stable under a set of actual and counterfactual *interventions*. For instance, Newton's law of gravity (i.e. the inverse square law) remains invariant under actual and counterfactual interventions, which would change the values of the masses of the gravitating bodies or the distance between them. So Woodward notes that "the notion of invariance is obviously a modal or counterfactual notion", since it has to do "with whether a relationship would remain stable if, perhaps contrary to actual fact, certain changes or interventions were to occur" (2000: 235).

Thinking about the difference between laws and accidents, we saw that laws support counterfactuals while accidents do not. Woodward shares the intuition behind this idea, but gives it a very different gloss. Not all counterfactuals are of the right sort for the evaluation of whether a generalization should count as a law. Only counterfactuals that are related to interventions can be of help. An intervention gives rise to an “active counterfactual”, that is, to a counterfactual whose antecedent is made true “by interventions” (1997: S31, 2000: 199). Take the generalization *All As are Bs*. An active counterfactual would describe interventions “that realise or bring about [the antecedent] A” (2000: 238). How does this help to distinguish between laws and accidents? Take the generalization *All coins in my pocket are dimes*, and consider the active counterfactual made possible by the following intervention: trying to slip a quarter into my pocket. Since the foregoing generalization fails to support the active counterfactual *if this quarter were slipped into my pocket, it would turn into a dime*, it does not remain invariant under interventions. Contrariwise, the generalization *All metals expand when heated* would count as a law since it supports the following active counterfactual: *if I were to heat this iron rod, it would expand*, and hence it would remain invariant under the relevant intervention.⁵

7.2.1 No laws in, no laws out

The use of counterfactual and causal notions (such as *intervention*) is not necessarily a problem for Woodward’s account, as he does not attempt to provide an analysis of laws in non-nomic, or non-causal, terms. As noted above, his view is not reductive. But be that as it may, there might be some cause for concern. In checking whether a generalization, or any relationship between magnitudes or variables, is invariant, we need to subject it to some variations/changes/interventions. What changes will it be subjected to? Those that are permitted, or are permissible, by the prevailing laws of nature. So suppose that we test the law of the pendulum to see whether it is a law or an accident. Suppose also that one of the changes envisaged is to see whether it would remain invariant if the experiment was made on a spaceship which moved faster than light. This, of course, cannot be done, because it is a *law* that nothing travels faster than

light. So, some *laws* must be in place before, based on considerations of invariance, it is established that some generalization is invariant under some interventions. Hence, the notion of “invariance under interventions” (Woodward 2000: 206) cannot offer an adequate analysis of lawhood, since laws are required to determine what interventions are possible. Couldn’t Woodward say that even basic laws – those that determine what interventions and changes are possible – express no more than relations of invariance? Take, once more, the law that nothing travels faster than light. Can the fact that it is a law be the result of subjecting it to interventions and changes? It’s not clear that it can. For it itself establishes the *limits* of possible interventions and control.⁶ I do not doubt that it may well be the case that genuine laws express relations of invariance. But this is not the issue, for the manifestation of invariance might well be the *symptom* of a law, without being constitutive of it.

It seems that Woodward must be committed to this symptom/constitution distinction. As he explains in detail, invariance does not characterize laws only; other relationships or generalizations, which cannot be deemed laws, display invariance, especially in the special sciences. For instance, Woodward notes:

There are generalisations that are invariant and that can be used to answer a range of what-if-things-had-been-different questions and that hence are explanatory, even though we may not wish to regard them as laws and even though they lack many of the features traditionally assigned to laws by philosophers. (2000: 214)

Note, however, that accidents do possess *some* range of invariance. For instance, the accidental generalization *All taxis in Athens are yellow* is invariant under a certain range of interventions, since if you apply for a taxi licence in Athens, your cab would have to be yellow. So, if invariance is to be found in laws as well as in non-laws, it should be at best a *symptom* of lawhood. What, then, does lawhood consist in? Now Woodward is perfectly happy with the thought that laws are not what philosophers have taken them to be. He (2000: 222) thinks that most of the standard criteria “are not helpful either for understanding what is distinctive about laws of nature or for understanding the features that characterise explanatory generalisations in

the special sciences". In particular, he takes it that in so far as a generalization is invariant under a certain range of interventions, it can be a law without being exceptionless (cf. 2000: 227–8). But no clear picture emerges as to what exactly makes a generalization a law. For, as Woodward (2000: 227) admits, even laws will *not* be invariant under *all* actual and possible interventions. For instance, Maxwell's laws break down at the Planck scale, where quantum mechanical effects take over. As a result of all this, the difference between (a) laws, (b) invariant generalizations that are explanatorily useful but non-laws, and (c) mere accidents, is deemed to be a difference "in degree . . . rather than of kind" (2000: 241). It is a difference in degree precisely because the notion of invariance under interventions admits of degrees. Some generalizations have a wider range of invariance, whereas others have a narrower range and yet others are "highly non-invariant" (2000: 237). This is not to say, Woodward claims, that the difference in degree is no difference at all. For, as he says:

the features possessed by generalisations, like Maxwell's equations [which are paradigmatic cases of laws] – greater scope and invariance under larger, more clearly defined, and important classes of interventions and changes – represent just the sort of generality and unconditionality standardly associated with laws of nature. (2000: 242)

Be that as it may, it should be stressed that laws are required in order to fix the range of invariance of a generalization. For, in order to specify the range of invariance of a generalization, we first need (a) to specify what interventions are physically possible, and (b) which of them, if they happened, would leave the given generalization unchanged. Both of the above, however, need a prior reliance on *laws*. As noted above, it is laws that specify the physically possible interventions. What needs to be added here is that it is laws that govern the assessment of the counterfactual in (b). For instance, specifying what interventions, had they happened, would have left Kepler's law unchanged requires holding other *laws* fixed. For if laws, for example Newton's laws, were allowed to be violated, then the range of invariance of Kepler's laws would be very limited. So it seems that Woodward's account boils down to the following

circular statement: a generalization is a *law* if it is invariant “under a large and important set of changes” (2000: 241), where the relevant set of changes is determined by *laws*.

This is as good a place as any to note that, in an earlier piece, Woodward (1992) tried to offer a metaphysical back-up to his views about invariance by accepting the existence of (objective) capacities. We shall postpone an examination of the nature of capacities until section 7.4. For the time being, we shall look into a different attempt to tie laws to invariance, that of Peter Menzies (1993).⁷

7.3 Agency

Where Woodward tries to base his account of laws on the metaphysical notion of capacity, Peter Menzies (1993) tries to offer a similar account based on the notion of human agency. He also starts with the thought that “a law of nature, but not an accidental regularity, is robust or *resilient under actual and hypothetical experimentation*” (1993: 207), but he adds that this notion of resiliency (which is akin to invariance) should be cashed out in terms of “a modal concept that we all possess in virtue of being decision makers” (*ibid.*). This is the concept of “a possible course of events within one’s control” (*ibid.*) The intuitive idea is quite clear. For a possible course of events to be within one’s control, it is enough that one be able to anticipate (or bring about by some actions) what might (or will) happen, if this course is followed. So one possible course of events within control of the reader of this book is that the reader, at time t , reads the book; the reader at time $t+1$ gets disappointed in it; then, at time $t+2$, decides that reading it is a waste of time; then at time $t+3$, decides to sell the book right away, and so on. Menzies, however, does not offer an exact definition of a possible course of events within one’s control. He takes it to be a *conceptual primitive*, but also clear enough to anyone who is a decision-maker, that is, practically, to everybody.

When it comes to the characterization of laws, Menzies talks of an “experimentally possible course of events” in order to broaden the agent’s arsenal. But the main thought remains the same. If a generalization is so resilient that there is no experimentally possible course of events in which it can break down, the generalization is

experimentally necessary. Hence, it is a *law*. Conversely, if we were dealing with an accident, then *some* experimentally possible course of events would violate it. Invariance, or stability, under experimentally possible courses of events is, then, the hallmark of a law. Menzies (1993: 209) defines the locution *It is a law that* explicitly:

It is a law that all *F*s are *G*s if and only if (i) it is experimentally necessary that all *F*s are *G*s and (ii) it is not logically necessary that all *F*s are *G*s.

Clause (ii) of the definition is necessary in order to distinguish experimental necessity, that is, the necessity by which laws hold, from logical necessity. Clause (i) suggests that Menzies takes laws to issue in exceptionless regularities. For if they didn't, then there would be experimentally possible courses of events that would violate them. Yet it is well known that some laws hold only under certain conditions. For instance, metals expand when heated only if the pressure is constant. Or the laws of ideal gases hold only if we disregard the molecular structure of the gases. So there are experimentally possible course of events that would lead to the violation of these laws. On Menzies' account then, they should not count as laws; hence, they should be deemed accidents. But this is hard to swallow. Even if there are experimentally possible courses of events that violate both the generalization *All metals expand when heated* and the generalization *All coins in my pocket are dimes*, there are still grounds to call the first a law and the second an accident. We could, perhaps, regiment the generalization that all metals expand when heated by including in it a number of clauses that make sure that it is not violated. But we could do exactly the same with typical accidental generalizations. So there is suspicion that Menzies' account does not adequately capture the distinction between laws and accidents.

Note also that according to Menzies (1993: 210), his theory suggests that being a law is an intrinsic property of a generalization: "it does not depend on complicated extrinsic considerations regarding its relations with other regularities". This is not quite right, however. For in order to specify the experimentally possible courses of events, we need to rely on laws: it is laws that dictate what is experimentally possible and what is not. So Menzies' account makes

lawhood not independent of the truth of those *laws* that determine what is experimentally possible. Besides, as Armstrong (1993c: 231) observes, since the notion of experimental intervention is a *causal* concept, laws will get back in through the window, for unless Menzies is committed to genuine singular causation, the experimentally possible courses of events will be lawlike. So far from offering an analysis of lawhood, Menzies has presupposed it in his own analysis, rendering the latter circular. Can, then, Menzies claim that the experimentally possible courses of events are instances of singular causings? Hardly, if he wants to make sense of the fact that they can be repetitively used in the testing of a law.

Menzies does anticipate and try to solve a number of objections to his account. But there is one, which he also considers, that cannot be easily laid to rest: his account seems to be too subjective to be able to offer a well-founded characterization of laws. Woodward tried to avoid subjectivity in his own similar characterization of laws by introducing *capacities*. But Menzies has nothing like this to ground the experimentally possible course(s) of events. Even if he had, the reference to agency, which produces the experimentally possible courses of events, seems ineliminable. In trying to meet this objection, Menzies attempts to analyse the concept of law along the lines of dispositional analyses of secondary-quality concepts (e.g. the concept of red). Let us not go into the interesting details of this attempt (but see section 3.4.1), for his conclusion is that if we follow his analysis the concept of law won't be "excessively subjective" (1993: 221). All the same, this conclusion is no consolation to those who think that there is nothing subjective in laws. Armstrong's (1993c: 231) remark seems very apt: "I find the demotion of a law to the status of secondary quality a trifle distressing – my own instinct is to make laws completely objective – but I don't know that I have an argument to back up my instinct."

7.4 Capacities to the rescue

As noted in section 7.2, Woodward supported his suggested difference between laws and accidents by saying that the truth-makers of laws are facts about "capacities, powers or dispositions of particular objects and systems" (1992: 196). It is, for instance, in virtue of the fact that the rod of the pendulum has a certain capacity to affect the

period of the pendulum that the equation $T = 2\pi\sqrt{l/g}$ expresses a law. Unlike traditional universals, Woodward's capacities are "firmly rooted in particular objects and systems, and do not exist apart from these" (*ibid.*). This suggestion makes Woodward's view different from Tooley's, who believes in the existence of uninstantiated universals, but one may wonder how it is different from Armstrong's, who takes universals to exist wholly in their instances. Take a metal rod that does not expand when it is heated, because an extreme pressure is exerted on both of its ends. Armstrong would say that the universal of expansion was not instantiated in this rod because it was prevented by the instantiation of another universal (extreme pressure on the rod's ends). Woodward would say that the rod's capacity to expand under heating was not manifested, because the manifestation of another capacity prevented it. One might wonder what the substantial difference between these two views is. To be sure, Woodward suggests that unlike universals, capacities might be possessed or exhibited only in a certain *range* of circumstances, and not outside them. So, capacities leave room for local laws, or laws that hold only in limited domains, or inexact laws, or laws that have exceptions. So far, so good. But one may still wonder (a) why we should call all these imperfect generalizations laws, and (b) how an appeal to capacities that are not uniformly manifested is anything other than an *ad hoc* move. Woodward thinks that it is an advantage of his appeal to capacities that it makes sense of the existence of uninstantiated (or vacuous) laws. So, a law may not have instances because the relevant capacities are *never* manifested. Couldn't we then say of any false generalization – for instance, that bodies rise if they are left unsupported – that the bodies involved in it have the relevant capacity to rise if they are left unsupported, although it is never manifested? In other words, what distinguishes unmanifestable capacities from non-existent ones?

7.4.1 Nomological machines

Recently, the view that capacities are prior to laws has also been advanced and defended by Cartwright (1989, 1999). She starts with the thought that no laws are strictly universal and exceptionless: all laws, even the supposed basic and fundamental laws of

physics, are *ceteris paribus*. That is, they hold only under certain circumstances, when *other things are equal*. For instance, Newton's first law, that all bodies retain their state of motion, holds under the circumstance that no forces are exerted on them. So Cartwright challenges the Humean view that laws are exceptionless regularities, since, she says, there are no such things.⁸ How then does it appear that there *are* regularities in nature, for example, that all planets move in ellipses? She does not deny that there can be regular behaviour in nature. But she claims that where there is regular behaviour in nature, there is a *nomological machine* that makes it possible. A "nomological machine" is "a fixed (enough) arrangement of components, or factors, with stable (enough) capacities that in the right sort of stable (enough) environment will, with repeated operation, give rise to the kind of regular behaviour that we represent in our scientific laws" (1999: 50).

Nomological machines make sure that "all other things are equal". So they secure the absence of factors, which, were they present, would block the manifestation of a regularity. Take, for instance, Kepler's law that all planets move in ellipses. This is not a strictly universal and unconditional law. Planets do (approximately) describe ellipses, if we neglect the gravitational pull that is exerted upon them by the other planets, as well as by other bodies in the universe. So the proper formulation of the law, Cartwright argues, is: *ceteris paribus*, all planets move in ellipses. Now, suppose that the planetary system is a stable enough nomological machine. Suppose, in particular, that as a matter of fact, the planetary system is (for all practical purposes) shielded: it is sufficiently isolated from other bodies in the universe, and the pull that the planets exert on each other is negligible. Under these circumstances, we can leave behind the *ceteris paribus* clause, and simply say that all planets move in ellipses. But the regularity holds only so long as the nomological machine backs it up. If the nomological machine were to fail, so would the regularity. As Cartwright has put it: "laws of nature (in this necessary regular association sense of 'law') hold only *ceteris paribus* – they hold only relative to the successful repeated operation of a nomological machine" (1999: 49–50).

Nomological machines might occur naturally in nature. The planetary system, for instance, is such a natural nomological machine. But, according to Cartwright, this is exceptional. As she says: "more often

[the nomological machines] are engineered by us, as in a laboratory experiment" (1999: 49). "In any case," she adds, "it takes what I call a nomological machine to get a law of nature" (*ibid.*).

For the operation of a nomological machine, it is not enough to have a stable (and shielded) arrangement of components in place. It is not enough, for instance, to have the sun, the planets and the gravitational force in place in order for the planetary machine to run. Cartwright insists that it is the *capacities* that the components of the machine have that generate regular behaviour. For instance, "a force has the capacity to change the state of motion of a massive body" (1999: 51). Couldn't the nomological machine itself be taken to be a regularity? No, she answers: "the point is that the fundamental facts about nature that ensure that regularities can obtain are not again themselves regularities. They are facts about what things can do" (1995a: 156). But what exactly are capacities, that is, the things that things can do?

7.4.2 What are capacities?

Cartwright (1989: 9) focuses attention on "what capacities do and why we need them" and not on "what capacities are". But later she says: "It is capacities that are basic, and laws of nature obtain – to the extent that they do obtain – on account of the capacities" (1999: 49). Yet no clear picture emerges as to what capacities are. To be sure, we are given examples of capacities and of how they relate to laws: "I say that Newton's and Coulomb's principles describe the capacities to be moved and to produce a motion that a charged particle has, in the first case the capacity it has on account of its gravitational mass and in the second, on account of its charge" (1999: 65). If laws describe what the entities involved in them can do on account of their capacities, then these capacities should be individuated, and ascribed, to entities, independently of the lawlike behaviour of the latter. But it is not clear how this is done. It seems that far from being independent of laws, the property of, say, *charge* is posited and individuated by reference to the lawlike behaviour of certain types of objects: some attract each other, while others repel each other in a regular fashion. The former are said to have opposite charges, while the latter have similar charge. Cartwright (1999: 54–5) says: "The capacity is associated with a single feature –

charge – which can be ascribed to a body for a variety of reasons independent of its display of the capacity described in the related law.” Suppose this is true. Still, it does not imply that the capacity is not grounded in any regularities at all. Cartwright disagrees. She (1999: 72) claims that “capacity claims, about charge, say, are made true by facts about what it is in the nature of an object to do by virtue of being charged”. One would expect, then, an informative account of what it is in the nature of an object to do. Specifically, one would expect that the nature of an object would determine its capacities, and would delineate what this object can and cannot do. But Cartwright goes on to say: “There is no fact of the matter about what a system can do just by virtue of having a given capacity. What it does depends on its setting . . .” (1999: 73).

Why, then, should we bother to attribute capacities? We could just offer an open-ended list of the things that a system does when it is placed in several settings. If, at least, there was a fact of the matter as to what a system *can* do by virtue of having a given capacity, the capacity could be used (a) to predict what a system can or cannot do, and (b) to explain why it behaves the way it does. Cartwright’s wording, however, is careful. It does not imply that there is no fact of the matter about what a system (or an object) can do by virtue of its *nature*. Yet one would expect that if the nature of an object placed some substantive constraints on its capacities, there would be a fact of the matter about what this object can do by virtue of its capacities. For instance, one would expect that although a certain particle has the capacity to move, its nature constrains this capacity so that it cannot move with velocity greater than the velocity of light. In any case, Cartwright is in need of a more detailed account of how capacities are individuated. It’s not even clear whether capacities are properties of *objects* (e.g. particles). Or are they properties of properties of objects (e.g. a property of the charge of the particle)? Or are they properties of properties *simpliciter* (e.g. a property of charge)? Here is a list of things attributed to capacities, according to Cartwright (1999):

- Capacities are not restricted to any single kind of manifestation (59).
- They are not to be identified with any particular manifestations (64).

- They have endless manifestations of endless different varieties (64).
- They are highly generic: they issue in a variety of different kinds of behaviour (54)
- They can be exercised more or less (67).
- They can combine with other capacities (54).
- They can be impeded by other capacities (71).
- They are open-ended (59).
- What makes capacity claims true are facts about capacities (72).

This list certainly constrains the nature of capacities. But it is far from offering a substantial understanding of what capacities are.

We noted above that according to Cartwright capacities need not issue in regularities. They may do so, if there is a nomological machine in place. But capacities, she insists, might well manifest themselves in *single* instances. When, for instance, we say that aspirins relieve headaches, we do not mean that aspirins always (or, even, more often than not) relieve headaches. We mean that aspirins have the stable capacity to relieve headaches. This capacity will be manifested only if the circumstances are right. This capacity, Cartwright (1989: 3) argues, would be there, even if there were just one *single* manifestation of relief from a headache, after taking aspirin. All this is important for Cartwright because she holds all of the following three theses:

- (a) causation is tied to the manifestation of capacities;
- (b) there is singular causation;
- (c) causal laws are, at best, generalizations over singular causings.⁹

If, then, capacities were manifested only via regularities, or if ascription of capacities was legitimate only if a regularity was present, then (b) would be in jeopardy. But there are two more theses that Cartwright also endorses:

- (d) the causal claims made in science involve ascription of capacities;
- (e) scientific hypotheses should be testable, and the tests should be reliable.

Since one of Cartwright's central claims is that capacities are measurable, there is no conflict between (d) and (e). There is,

however, a residual problem. If the capacity of a factor x to bring about y was manifested regularly, then one could say that the presence of the capacity can be tested. Hence, one could move on to legitimately attribute this capacity to x . But if a capacity can manifest itself in a *single* case, then it is not clear at all how the presence of the capacity can be tested. Why, in other words, should we attribute to x the capacity to bring about y , instead of claiming that the occurrence of y was a matter of chance? So, there seems to be a tension between Cartwright's claim that capacities are testable and her claim that capacities are manifestable even in single cases.

Besides, as Margaret Morrison (1995) has noted, Cartwright seems to face another important problem. Do claims about the presence of capacities have extra content over the claims made by ordinary causal laws? Take, for instance, the ordinary causal law that aspirin relieves headaches. If we ascribe to aspirin a *capacity* to relieve headaches, would we gain in content? There is a sense in which we would. Ordinary causal laws are *ceteris paribus*, whereas capacity claims are not. Since it is only under certain circumstances that aspirin relieves headaches, it is only *ceteris paribus* true that aspirin causes headache relief. But once it is established that aspirin has the *capacity* to relieve headaches, this last claim is strictly universal: the capacity is *always* there, even if there may be contravening factors that block the manifestation of this capacity in some cases. The problem with this attempt to substantiate capacities is that the strictly universal character of claims about capacities *cannot* be established. If it is allowed that claims about the presence of capacities might be based on single manifestations, then it is not quite clear what kind of *inference* is involved in the movement from a single manifestation to the presence of the capacity. If, on the other hand, it is said that claims about capacities are established by ordinary inductive methods, based on several manifestations of the relevant capacity, then all that can be established is a *ceteris paribus* law. Based on cases of uses of aspirin, for instance, all that can be established is that *ceteris paribus*, aspirin relieves headaches. So it is questionable that talk about capacities has extra content over talk about ordinary causal laws.

Cartwright could argue that claims about capacities are strictly universal in the sense that objects have capacities even if they *fail* to manifest themselves. However, she would then seem to compromise

her view that capacities are measurable and testable. If unmanifestable capacities were allowed, then, as we have already seen when we discussed Woodward's views (see towards the end of section 7.4), the attribution of capacities would be in danger of trivialization. So, if Cartwright insists on single manifestation of capacities, she faces a sticky *trilemma*. Either talk of capacities does not have extra content over talk in terms of ordinary causal laws, or there is a mysterious method that goes from a single manifestation to the capacity, or there are unmanifestable capacities. All three options have unpalatable consequences.¹⁰

7.4.3 A metaphysical double vision?

Talk of capacities is not necessarily problematic. Yet it is not well grounded if it is not accompanied by a theory of what they are, how they are manifested, what individuates them, what governs their manifestation, and so on. Not without good reason, some philosophers (e.g. Mackie 1977, Earman 1984) take talk of capacities, or dispositions, to be either metaphysically innocuous or deeply problematic. Suppose that one wants to take capacities seriously. Then there are two ways to fix their ontological status (cf. Mackie 1977: 365–6). One, which Mackie calls “realist”, is to think that capacities (or dispositions) require a non-dispositional (categorical) ground. When an object manifests a capacity (disposition), when, for instance, a sugar cube put in water manifests its capacity to dissolve in water, then its relevant behaviour (its dissolution in water) is caused by some non-dispositional/categorical property of the object (e.g. the sugar cube's molecular structure). On this realist view, talk of capacities (or dispositions) is legitimate precisely because it is grounded in categorical properties of the things to which the capacities are attributed.¹¹ In particular, the attribution of a capacity to an entity is not an addition to the attribution to it of the categorical property. Rather, if this categorical property is not known, the capacity is a place-holder for the categorical property which, were it to become known, could explain fully (i.e. without further reference to capacities) the relevant causal behaviour of the object.

The other way to think of capacities is called “rationalist” by Mackie.¹² On this view, capacities (powers/dispositions) are

intrinsic properties of things “whose essential nature it is to tend towards the corresponding manifestation” (Mackie 1977: 366) and which are such that they entail that if the object were to manifest its capacity, a certain behaviour (or effect) would ensue. So, for instance, the capacity (or power/disposition) of sugar to dissolve in water is, on the rationalist view, an intrinsic property of sugar, suitably distinct from its molecular structure. Yet there are a number of reasons to reject the “rationalist” view. Mackie (*ibid.*) offers a host of them, but perhaps the most important ones are the following two. First, intrinsic capacities (powers) are redundant. Intrinsic categorical properties are enough to account for the causal effects of the supposed exercise of the relevant capacities. Secondly, positing intrinsic capacities seems to be the result of “metaphysical double vision: they just *are* the causal processes which they are supposed to explain seen over again as somehow latent in the things that enter into these processes” (*ibid.*).

In fact, if we start this double counting, there is no end to it. For apart from accepting that, say, copper has a certain molecular structure that accounts for its causal behaviour, we should accept that – on top of it – it has the capacity (power) to conduct electricity, another capacity to conduct heat, yet another capacity to melt at a certain temperature, and so on. All this is just “gratuitous multiplication” (*ibid.*). These thoughts are relevant to both Cartwright’s and Woodward’s projects, since it’s not clear what stance they take on the nature of capacities. They seem to take a “rationalist” stance, for to say the least, the “realist” view is consistent with the Humean approach to laws they deny. Understood in the “realist” way, capacities and the cognate do not hold, as Earman (1984: 203) claims, “non-Humean powers”.

All this, however, is very far from being the final word on dispositions/capacities. A lot of recent work has cast new light on the nature of dispositions (cf. Mumford 1998, Mellor 2000, Ellis 2000). In particular, there may be ways to hold on to a “realist” view on dispositions, while denying that each and every disposition has a categorical base. In his enlightening 1998 book, Stephen Mumford argues for “a functionalist theory of dispositions where the dispositional and the non-dispositional are understood as two distinct ways of denoting the very same instantiations of properties” (1998: 23).¹³

7.5 Normativity and stability

As we have already seen, RVL is committed to the following thesis: the (true) statement “It is a law that all *F*s are *G*s” and the (true) statement “All *F*s are *G*s” express the very same fact; namely, a certain Humean regularity. According to the different versions of RVL, the added locution *It is a law that* does indeed flag a different *attitude* towards the truth expressed by the statement “All *F*s are *G*s”. For instance, to say that “It is a law that all *F*s are *G*s” is to say that the generalization “All *F*s are *G*s” can be used in explanation and prediction. Or to say that “It is a law that All *F*s are *G*s” is to say that the generalization “All *F*s are *G*s” is part of the best deductive system. But, and this is the crux, the added locution does *not* signal that the fact expressed by the (true) statement “It is a law that all *F*s are *G*s” has a different (or a special) metaphysical status.

In his recent book, Lange (2000) has aimed to offer a new challenge to RVL (as well as to the alternative ADT approach). Briefly put, Lange aims to offer a *functional* account of lawhood: the laws of nature (i.e. the nomic facts) are characterized (and individuated) by reference to the distinctive role they play in the logic of scientific reasoning. As he (2000: 30) puts it, what he is looking for is “what a natural law must be in order for it to function as scientific reasoning presumes”. And he adds: “My approach to understanding lawhood is through understanding the special roles performed in scientific practice by claims believed to express laws” (*ibid.*).

7.5.1 The root commitment

Very much like the advocates of RVL, Lange too (2000: Ch. 1) is rightly impressed by the fact that laws function in a certain way in science. They are intuitively different from accidents, they support counterfactuals, they are essential to scientific explanation, they issue in predictions, their positive instances confirm them. But unlike with RVL, Lange thinks there is no simple characterization of the distinction between laws and accidents based on *any* of the above features of laws. For instance, he persuasively argues that there can be accidentally true generalizations, such as *All coins in my pocket are dimes*, which nonetheless support relevant counterfactuals – if, for instance, I had a strict and inviolable policy such that if any coin were to end up in my pocket it would have to be a dime.

However, where the advocates of RVL thought that, barring our distinct epistemic attitudes, the expression “It’s a law that p ” expresses no further fact beyond the corresponding Humean regularity, Lange thinks that it does express “a certain fact that goes beyond any corresponding Humean regularity” (2000: 24). This excess content of talk about laws captures, we are told, “objective” facts, that is, “facts that [do] not depend on our beliefs”, since “we could be mistaken about [them]” (*ibid.*). Yet, in a rather puzzling move, Lange identifies this excess content of laws over Humean regularities with something “irreducibly normative”: “a fact about how . . . we *ought* to reason” (*ibid.*).¹⁴ The extra *normative* content of a law is said to be associated with a *rule of inference* (2000: 24). So when it is believed that it is a law that all F s are G s, then an inference from “ a is F ” to “ a is G ”, made in accordance with the claim “All F s are G s”, is legitimate and acceptable.

Lange’s view might remind the reader of the inference-ticket approach discussed in section 7.1. However, Lange is no advocate of this approach.¹⁵ Part of his distance from the traditional view is accounted for by what Lange has called the “root commitment” we undertake when we believe that a certain claim p expresses a law: when a claim p is believed to express a law, it is not just believed to be true (or, at least, reliable); more importantly, it is believed to have been confirmed “in a special manner” (2000: 8). This special manner is what Lange calls “inductive confirmation” (more on this later). Its crux is that when a claim p is confirmed inductively it is believed also to hold for a range of *counterfactual* suppositions. From this “root commitment”, Lange argues, it follows “that p can perform all of the special functions that set the laws apart” (2000: 8). In particular, it follows that it can function as a rule of inference, that is, that it can be employed, “in certain circumstances”, as an acceptable step “in the course of making certain kinds of empirical predictions” (2000: 24).

We have been told that laws have a normative dimension. But where exactly is this located? I think that it cannot be in the laws themselves. Hooke’s law – that the restoring force F exerted by a spring under tension is proportional to its displacement x from its equilibrium length – does not have any normative content. It merely *describes* how springs behave. (Since it is the law statement and not the law itself that describes how springs behave, it’s more

correct to say that Hooke's law simply *governs* the behaviour of springs.) Kepler's first law, to use another example, does not say that planets ought to move in ellipses; it merely governs their motion. *Normativity*, to use Sellars's expression, lies in the space of reasons and not in the space of laws. These, perhaps tedious, points are meant to motivate the thought that the extra normative content that Lange has tried to identify does *not* belong to the laws themselves, but rather to our *beliefs* about them, and in particular to what we ought to do with statements *believed* to express laws. So Lange must be seen as suggesting that there is (must be) a normative surplus in our belief that a certain statement *p* expresses a *law*: when we believe that "it is a law that *p*" we are *entitled* to employ the statement *p* in a certain way; namely, as a step in our reasonings about matters of fact, as a premise in predictive arguments, as a means to support certain counterfactual assertions, and so on.

Once we have drawn, as we should, a distinction between the laws themselves and our *beliefs* about them, then it's no longer obvious that an appeal to the normativity of law statements can offer a solid ground to characterize what laws of nature *are*. Three points seem relevant here. First, as Lange himself notes, the normative component associated with a law statement is not absolute: it gives us an inferential licence only "in certain circumstances" (2000: 25). For instance, one can infer the value of the displacement of a certain spring under a certain force only if certain idealizations are accepted. But then the normative force of a law statement is limited. It does not tell us how we ought to reason; it only tells us how we ought to reason, all else being equal. Secondly, although Lange accepts that the notion of normativity associated with a law is weak, in the sense that it amounts to an *entitlement* to infer in accordance with the law and not to a *compulsion* to infer in accordance with it, this weak notion of entitlement seems equally applicable to accidents. Accidentally true generalizations might well entitle us to employ them in certain inferences. For instance, to use Lange's (2000: 13) own example, the accidental generalization *All pears on the tree are ripe* might legitimately be used as a step in a reasoning process concerning the status of a pear in a remote branch of the tree. Thirdly, it is not clear why this weak notion of normativity confers on the belief that a statement *p* expresses a *law* a superior status over the belief that *p* is (simply) true. The belief

that a statement is true (or the belief that a certain statement expresses a regularity) might be *enough* to account for the special role of this statement in reasoning. If it's true that nothing travels faster than light, then why isn't its *truth* enough to account for its special role in science? Why should this truth be deemed a law?

7.5.2 Preservation of laws

In any case, Lange does offer a more robust characterization of laws, based on the thought that laws govern not just what actually happens but also what would or could have happened under various possible circumstances. Specifically, he builds an account of laws in relation to their ability to support *counterfactuals*. We have already seen in section 7.2 that Woodward (among other current philosophers) has taken the prime characteristic of a law to be its invariance under a certain range of counterfactual suppositions. Now, although this is a thought also shared by Lange, he correctly notes that it's too quick to serve as an adequate characterization of lawhood (cf. 2000: 13–15). Accidents may also remain invariant under *some* range of counterfactual suppositions. So Lange modifies the notion of invariance in order to be able to characterize laws and to distinguish them from accidents. In what follows, I shall first try to sketch the basic elements of Lange's account and then offer a few critical points.

There is an intuitive difference between laws and initial conditions. Had the acceleration of a body been different, so too would be the force exerted on it. But Newton's second law would not thereby fail to hold: it remains invariant. So one might try the thought that laws are invariant under *all* physically possible changes in the initial conditions (actual and counterfactual ones). Carroll (1994: 59, 182–9) did, in fact, propound a related view. According to him, to say that it is a law that L is to say that, for any physically possible situation p , L would (still) be a law if p were the case. Lange too works with a similar idea, what he calls “preservation” of laws (2000: 47–55). Roughly, the idea is that a law L is preserved (i.e. it would still have remained true) under *all* physically possible counterfactual suppositions. The issue then is to specify what a physically possible counterfactual supposition is. Note that some counterfactual suppositions cannot be envisaged

unless some laws are violated. For some counterfactual suppositions are *inconsistent* with laws. For instance, the counterfactual supposition according to which a body could have had velocity greater than the speed of light cannot be envisaged unless the law that nothing travels faster than light is violated. Consequently, there must be a *restriction* to the possible characterization of laws in terms of invariance (preservation) under actual and counterfactual suppositions: the law must remain invariant under *all* (actual and counterfactual) suppositions *that are consistent with laws*. After all, it is the laws that dictate the *physically possible* suppositions.

The upshot of all this is twofold. First, the characterization of laws in terms of preservation (or invariance) under all physically possible counterfactual suppositions must be *collective*. Each law, taken individually, will fail to be preserved under *some* counterfactual suppositions, unless these suppositions are disallowed because they are inconsistent with all laws. Only all laws, taken collectively, are guaranteed preservation under all counterfactual suppositions, since they (collectively) dictate what is physically possible and what is not. Secondly, it seems to become inevitable that the characterization of laws in terms of preservation (invariance) is *circular*. If a fact *L* is a *law* if and only if it is preserved under all counterfactual suppositions that are logically consistent with *laws*, then the circle could not be more obvious.

Lange does believe that there is a kernel of truth behind the idea of “preservation” and aims to extract it in a way that is not vulnerable to the charge of circularity. To this end, he introduces the all-important notion of “stability” (2000: 99ff). Roughly put, a set of truths is stable if and only if its members are preserved under all suppositions that are consistent with each of them. Let’s say that a proposition *m* is *preserved* under a supposition *r* if and only if the following holds: if *r* had been the case, then *m* would have (still) been the case. To illustrate the concept of stability, Lange invites us to consider the following two sets: the set *T* of all truths expressed in the language of science and the set *LT* of all logical truths. Both sets are “stable” relative to all propositions that are consistent with each of their members (2000: 100). For instance, for any logical truth *m*, *m* is preserved by (since it logically follows from) every supposition *r*, which is consistent with the logical truths. Lange’s important contribution to the debate about laws (2000: 101–3) is

the proof of the following theorem: there is a *unique* set of truths *NT* which possesses non-trivial stability relative to all suppositions logically consistent with each of its members.¹⁶ *NT*, Lange contends, can be seen as the set of *laws* – collectively characterized. It's *not* the only set with stability (since, as we have just seen, both the set *T* of all truths *and* the set *LT* of all logical truths are stable); but *NT* is the only set with *non-trivial stability*. This set *NT*, which might be called the set of nomological truths, can capture a sense of necessity that is intermediate between the logical (or conceptual) necessity of *LT* and the pure contingency of *T*.

7.5.3 What laws are vs how we get at them

We need not go into the details of Lange's proof. Suffice it to make a number of observations about the concept of *stability*. First, as was noted above, attempts to show that laws display invariance under counterfactual suppositions may well lead to a rather tight circle. If a fact *L* is a *law* if and only if it is preserved under all counterfactual suppositions that are logically consistent with *laws*, then we need to know what the laws are in order to specify what the laws are. Lange's suggestion avoids the charge of circularity rather nicely. For the stability of the set *NT* is not defined with reference to laws; it is fixed by the members of *NT* themselves – whatever they are. It is then a further step to suggest that this unique set – with the property of non-trivial stability – is the set of *laws*.

A natural worry, however, is how it is ensured that this set contains no accidents. This is especially pressing for Lange, since he wants to argue that laws differ from accidents “not merely in degree but in kind” (2000: 105). Although we cannot go into all the relevant details here, it is noteworthy that Lange tries to show that if some accidents were included in the set *NT*, then this set would no longer be stable. So the set of laws differs “in kind” from any set containing accidents, in that the set of laws alone is non-trivially stable. The intuition behind Lange's argument is quite clear. If we were to add to the set *NT* of laws the accidental generalization *All coins in my pocket are dimes*, then this generalization would not be preserved under all counterfactual suppositions that are consistent with the members of *NT*. For instance, it wouldn't be preserved under the following counterfactual supposition: if I went to a

country in which all currency was in coins, and none of them were dimes, then all coins in my pocket would *not* be dimes. So if an accident were to be added to *NT*, this would no longer be stable. This is all fine, but one might worry about the following. Instead of underwriting the distinction between laws and accidents, Lange's suggestion seems to require that such a distinction is *already* in place: it is because we already accept that the generalization *All coins in my pocket are dimes* is accidental and not lawlike that we take its addition to *NT* to violate *NT*'s stability.

A second (related) observation about Lange's account is that although it identifies the extension of lawhood it seems to fail to specify its intension. As we have already seen, the members of *NT* are identified collectively. Besides, we are given a criterion for *NT*-membership, that is, *stability*. But even if this criterion were enough to separate laws from accidents, we are still *not* told what laws are, that is, what exactly this property of lawlikeness that some truths have while others lack consists in. We are told that laws are stable, but what exactly is the thing we attribute stability to? It might be objected that this last point is overstated. For Lange aims primarily to show how laws *function* in science and not what laws are. Still, it would be important to know what exactly is attributed to a truth when it is deemed to be a member of the only non-trivial stable set *NT*.

A third observation is that, as Lange himself is fully aware, the notion of stability is not enough to determine how to tell whether a certain fact is a law. In particular, the notion of stability does not yet tell us *why* we (better, the scientists) should be interested in discovering whether a certain fact is nomic. In order to address this issue, Lange suggests that we "should turn to the relation between our beliefs about the laws and the inductive confirmations that we carry out" (2000: 110). His prime thought is that there is a link between *believing* that a fact is a law and being able to confirm inductively a hypothesis that expresses it.

As we saw in section 7.1, Cohen suggested that accidental generalizations are the products of enumerative induction, while lawlike ones are the products of the more sophisticated eliminative induction. In a rather interesting move, Lange tries to reinstate good old enumerative induction as *the* means to establish lawlike generalizations and to *inductively confirm* them. Yet Lange's concept of

“inductive confirmation” is novel in two respects. First, he takes it to be the case that when a hypothesis is confirmed inductively, it is said to apply not only to its actual unexamined instances, but also to certain of its *counterfactual* instances. Secondly, Lange takes it to be the case that the hallmark of inductive confirmation is that the very same *reason* that makes us project a hypothesis to unexamined instances also makes us project it to counterfactual instances, this reason being whatever evidence we have that the hypothesis describes a physical necessity (cf. 2000: Ch. 4). So it seems fair to say that the members of *NT* (that is, the laws) are taken by Lange to be (and hence to be identified with) the products of sound inductive reasoning – what, more broadly, Lange (2000: 143) calls “inductive strategies”. The upshot of an inductive strategy is that a hypothesis which is deemed a law is projected “indiscriminately” (2000: 148). This “indiscriminate” inductive projection, Lange argues, can account for the difference between laws and accidents.

Take the stock example: *All gold cubes are smaller than one cubic mile*. This, Lange (2000: 148) notes, is an accidental generalization because it cannot be projected indiscriminately: there is a range of *possible* circumstances in which it would not apply, for example, if a multibillionaire wanted to buy enough gold to actually build the envisaged gold cube. Contrariwise, the generalization *All plutonium cubes are smaller than one cubic mile* is lawlike because there is no possible situation in which it would fail to hold. In fact, Lange intends to push his claims about inductive confirmation one step further. He wants to argue that there is a link between the concept of stability and the concept of inductive confirmation: the range of possible situations to which the inductive confirmation of a set of hypotheses entitles us to project them indiscriminately *is* the range over which the laws (the hypotheses we should so project) are *stable*. Hence, as Lange (2000: 157) puts it, the inductive strategies are the best means to “identify the members of [NT]”, that is, the laws of nature. With this last move, Lange aims to deflect an obvious charge that can be levelled against his account of inductive confirmation; namely, that it ends up offering an *epistemic characterization* of the laws of nature. For, the charge will surely be, it seems that by insisting on identifying laws via the method (“inductive strategies”) by which they are arrived at and the manner in which they are confirmed, Lange does not improve significantly over the

epistemic version of RVL (see section 5.4). Yet by tying inductive confirmation to the notion of stability, he seems able to offer a more objective account of what laws of nature are. For, as he puts it, “what makes a set of inductive strategies best is, in part, the correctness of certain subjunctive conditionals” (2000: 156). As a result of this, lawhood acquires “an unreduced metaphysical (rather than epistemic) element” (*ibid.*).

These last remarks will lead us to a final general point concerning Lange’s views. Although counterfactuals feature prominently in Lange’s account of laws, we are not told when exactly a counterfactual is true (or correct). Indeed, Lange stresses that he offers “no account of what makes counterfactuals correct” (2000: 10), although he assumes that they have “some sort of objective correctness” (*ibid.*). The consequences of this are, perhaps, underestimated by Lange. He claims that, having specified the relation between laws and counterfactuals, he does not also need to “determine whether the laws are laws in virtue of certain counterfactuals holding or whether certain counterfactuals obtain partly in virtue of which facts are laws” (*ibid.*). Yet things are not so straightforward. On the one hand, the very concept of stability requires the assessment of several (occasionally very complicated) counterfactuals, such as if so-and-so had been the case, *L* would have still been a law.¹⁷ Unless the conditions under which they are correct are specified, one cannot offer a robust account of stability. Take, for instance, Kepler’s first law and wonder about the following counterfactual: had there been a very massive planet very close to Mars, all planets would still describe ellipses. Whether or not it is a law that *all planets move in ellipses* will depend on its being preserved under the foregoing counterfactual supposition, and finding this out will require assessing the foregoing (moderately complicated) counterfactual. If we are not told what makes counterfactuals such as the above correct (or incorrect), we are not told whether Kepler’s law is indeed a law. On the other hand, Lange’s very idea of a fact being an accident and not law depends on the assessment of some counterfactuals. Unless the conditions under which they are correct are specified, one cannot even start offering a robust account of what distinguishes between laws and accidents.

7.6 Properties and laws

Promising to offer a unified account of deterministic and indeterministic laws, Mellor (1991) suggests that all laws involve objective (single-case) chances and are embodied “in the actual properties and relations (including chances) they contain” (1991: xviii). Mellor’s views are quite complicated and deserve more careful study and thought. Before I try to unpack the seemingly obscure notion of *embodiment*, let me say a few things about Mellor’s position in the overall debate.

Mellor (1991: 168) rejects the view that laws are, essentially, regularities. He does agree with Armstrong’s realism about universals (cf. 1991: 170) and he does think that laws involve properties (universals): “the law that all *G*s are *H* involves the properties *G* and *H*” (1991: 155).¹⁸ Yet he takes it that the ADT view of laws, that is, the view that laws are necessitating relations among universals, is deeply flawed. He notes that the notion of necessitating relation is “*ad hoc* because there is nothing more to [it] than what [it is] defined to do, namely make laws necessary and enable them to support their counterfactual conditionals” (1991: 168). For him, laws of nature are not metaphysically necessary either. So they are contingent (1991: 145ff). Given all these strictures on the contours, what shape does his view of laws have?

For a start, Mellor takes chances seriously: chances, expressed by single-case probabilities, are objective features of the world. They are “tendencies” of actual particulars, and they exist as “real properties” which partake of causal give-and-takes, irrespective of whether determinism or indeterminism is true. If indeterminism is true, then chances are “real contingent possibilities” (1991: 158). For instance, the chance of a smoker getting lung cancer is “the degree of possibility” that the smoker will get it, contingent on his or her smoking. If determinism is true, then the chance of a smoker getting lung cancer is the “quantitative tendency” to develop lung cancer, contingent on his or her smoking and on whatever other factors (maybe genetic) lead to lung cancer. Let’s not dwell on this issue any more, except to stress that, for Mellor, both deterministic and statistical (indeterministic) laws involve chances. To say, for instance, that it is an indeterministic law that radioactive atoms of type *R* have a probability 0.5 to decay within time *t*, is to say that *R*-atoms have an (objective) chance 0.5 of decaying within time *t*.

Similarly, to say that it is a deterministic law that *All Fs are Gs* is to say that *Fs* have an (objective) chance 1 of being *G* (1991: 162). As he puts it: “deterministic laws are just limiting cases of indeterministic ones” (*ibid.*). To simplify matters, let’s concentrate on deterministic laws. Since he takes chances to be properties, Mellor says that a law of the form *All Fs are Gs*, which asserts that the chance of an *F*’s being *G* is 1,¹⁹ involves three properties: *F*, *G* and $C_1(G)$, where $C_1(G)$ is the *property of F-having-the-chance-1-of-being-G*. Then, he claims, the property $C_1(G)$ is enough to guarantee that it is a law that all *Fs* are *Gs*. Properties such as $C_1(G)$ already embody a “contingent law” that certain *Fs*, namely those with property $C_1(G)$, are all *Gs* (cf. 1991: 164). As he (1991: 166) says: “the law that all *Fs* are *G* contains only *F* and *G* (and $C_1(G)$, an addition I shall hereafter take as read), and the particulars that instantiate it are only *F* and *G* . . . [the dummy letters have been changed]”. Contrary to the ADT view of laws, Mellor thinks that *no further relation of necessitation* is needed between properties *F* and *G*.

Some more light might be cast on Mellor’s suggestion if we take into account his view of how properties are identified. For him, they are identified a posteriori by looking at scientific theories (cf. 1991: 175). He explains how this happens by appealing to a device known as a Ramsey-sentence. In order to get the Ramsey-sentence RT of a theory *T*, we replace all theoretical constants with distinct variables $\{u_i\}$, and then we bind these variables by placing an equal number of existential quantifiers $\exists u_i$ in front of the resulting formula. So, suppose that the theory *T* is represented as $T(t_1, \dots, t_n; o_1, \dots, o_m)$, where *T* is an $m+n$ -predicate. The Ramsey-sentence RT of *T* is: $\exists u_1 \exists u_2 \dots \exists u_n T(u_1, \dots, u_n; o_1, \dots, o_m)$. There is no need to go into much detail about the nature of Ramsey-sentences.²⁰ Nor, for our purposes, do we need to assume a distinction between theoretical and observational terms. All we need is the idea that when we replace a theory by its Ramsey-sentence, we are only committed to the existence of entities that make the theory true, without naming them.

Now Mellor says that the real properties are those properties that are (would be) “quantified over by the Ramsey sentence of the conjunction of all true law statements” (1991: xviii). So suppose that we have the conjunction of all the true law statements, and we

form their Ramsey-sentence Σ . The universals (properties and relations) whose existence is implied by Σ are the real properties: “there is no more to them than the totality of laws they embody” (1991: 167), a totality which is nothing other than Σ . Mellor takes it that Σ will entail law statements of the form “All F s are G s”. So Σ will entail that there are properties F and G such that $L(F, G)$, where L will be a *predicate* describing the relation between properties F and G . $L(F, G)$ looks very much like Armstrong’s necessitating relation $N(F, G)$ (see section 6.2). So why aren’t the two the same? Appearances to the contrary, Mellor says, L is just a predicate and *not* a universal. So it isn’t Armstrong’s necessitating relation.

All this will need further reflection (and study) to be taken in.²¹ So I won’t tire the reader with further details. I will only briefly raise three issues that need more attention. The first is whether Mellor’s account is circular. Properties, we are told, embody laws. But, at the same time, laws are what give universals their identity (cf. 1991: 161, 173). So we cannot have an account of laws without some account of properties (since laws are embodied in them). But we cannot have an account of properties without an account of laws, since we get at the properties *via* laws – and in particular via the Ramsey-sentence Σ of the conjunction of all law statements. Mellor’s thought seems to be that laws and properties are so interconnected that no account of each of them can be given without an account of the other. The second issue relates to the nature of Ramsey-sentences. One may wonder how we get from the fact that the existentially quantified variables of a Ramsey-sentence quantify over some entities to these entities being *universals*. And one might equally wonder whether there is a unique Ramsey-sentence of the conjunction of the law statements. Although Mellor (1991: 167, 175, 1995: 193–6) does deal with these issues, it is not straightforward that what we get out of a Ramsey-sentence are properties. The third and final issue is the relation between Mellor’s approach and the MRL view of laws. Taking the Ramsey-sentence Σ of the conjunction of all law statements is very close to taking the existentialised version of the MRL “best system”. Mellor (1991: 167–8) is fully aware of this connection. He notes: “. . . law statements would indeed be the general axioms and theorems of such a system” (1991: 168). So his disagreement with the MRL view comes to this. He takes it

that the statements that have a position in the *best system* are there *because* they express laws, and not that some statements express laws because they have a position in the *best system*. This sounds exactly right to me. This is how things should be. But it's not clear that Mellor's account has shown *why* this should be so. The problem raised under the first issue above is one reason why Mellor's entirely correct claim has not yet been fully established.

7.7 Taking stock

All theories we have examined in Part II of the book start with the assumption that there are laws of nature, and try to explain what they consist in.²² But, so far, we have not found a problem-free characterization. However, let us attempt a brief cost-benefit analysis. Let us start with the MRL view of laws. On the benefit side:

- Its metaphysics is unproblematic: laws are regularities.
- Its epistemology is clear: laws can be known by the ordinary inductive methods.
- It shows how facts about laws supervene on non-nomic facts.
- It is close to scientific practice, that is, to what scientists take laws to be.
- It is able to deflect some of the most serious objections levelled against the Humean approach to laws: it shows that there is a robust, if not entirely objective, distinction between laws and accidents; it can accommodate the existence of uninstantiated laws; it can show, to some extent, at least, how laws can support counterfactuals; it shows that what regularities constitute laws is not just a reflection of our subjective or epistemic preferences.

On the cost side:

- It seems to deprive laws of some, intuitively clear, sense of necessity.
- It does leave a residual quasi-subjective element to the characterization of lawlikeness.

Now, the ADT view. On the benefit side:

- It shows how laws are relations among properties.
- It makes lawhood entirely objective.
- It shows how uninstantiated laws can be accepted.
- It issues in an *objective* distinction between laws and accidents.
- It shows how laws can support counterfactuals.

On the cost side:

- Its metaphysics is problematic: the notion of necessitation remains unclear.
- Its epistemology is legitimate, since it is based on inference to the best explanation. Yet it is not clear that positing necessitating relations among universals is indeed the best explanation of lawhood.
- It's not in contact with the concept of a law of nature, as this is used in science.
- By denying that nomic facts supervene on non-nomic facts, it makes a mystery of how laws issue in regularities.

The alternative views of laws we saw in this chapter are too diverse to admit of a collective cost–benefit analysis. Some of them, for example, the approaches based on the notion of invariance, seem quite promising. Yet, none of them seems to be problem-free. So, on balance, we might well say that the least cost and the most benefit accrues to the MRL view. Many will deny this. Some will stress that the both the MRL and the ADT views are wrongly focused on the metaphysics of laws and that the focus should be shifted to methodological issues. Be that as it may, we shouldn't forget that the debate still goes on. So let's leave all this behind, and turn our attention to the concept of explanation.

Explanation

Deductive-nomological

8 explanation

8.1 The empiricist legacy

The modern empiricist approach to the connection between causation and explanation was shaped by Hume's critique of the relation between cause and effect. As was already noted in the Introduction, the logical empiricists took Hume to have offered a *reductive* account of causation, and in particular one that frees talk about causation from any commitments to a necessary link between cause and effect. Within science, Carnap stressed, "causality means nothing but a functional dependency of a certain sort" (1928: 264). The functional dependency is between two states of a system, and it can be called a "causal law" if the two states are in temporal proximity, and one precedes the other in time. Schlick expressed this idea succinctly by pointing out that:

the difference between a mere temporal sequence and a causal sequence is the regularity, the uniformity of the latter. If C is *regularly* followed by E, then C is the cause of E; if E only 'happens' to follow C now and then, the sequence is called mere chance. (1932: 239)

Any further attempt to show that there was a necessary "tie" between two causally connected events, or a "kind of glue" that holds them together, was taken to have been proved futile by Hume, who maintained that "it was impossible to discover any 'impression' of the causal nexus" (Schlick 1932: 246). The twist that logical empiricists gave to this Humean argument was based on

their verifiability criterion of meaning: attributing, and looking for, a “linkage” between two events would be tantamount to “committing a kind of nonsense” since all attempts to verify it would be necessarily futile (Schlick 1932: 245).

In fact, the concept of causation was taken to be a kind of test case for the positivists’ distinction between science and metaphysics. When we examine the notion of *causal relation*, Carnap (1928: 35–6) suggested, there are two problems that need to be distinguished: the “correlation problem” and the “essence problem”. The essence problem is metaphysical, Carnap thought, because it purports to investigate what the alleged essence of causation is, beyond regularity. As such, Carnap went on, it relies on the “erroneous assumption” that there is something in causation beyond correlation (“i.e. beyond mathematical function”). The correlation problem, on the other hand, is empirical. It investigates what events are correlated, where correlation is understood as subsumption under a “general functional law”. Carnap immediately added that the problem of correlation is none other than finding “the laws of nature”. But for him, these “laws of nature” are not causal, if by that we mean anything other than that they express a “functional dependency of a certain sort” between two states of a physical system, or between two physical magnitudes (1928: 264). To be sure, Carnap did not want to excise talk of “cause” and “effect” from science, although he certainly toyed with this idea. But he insisted that the only meaningful content this talk can have is when we call “cause” the event, or the physical magnitude, or the physical state, which temporally precedes another one nomologically dependent on the former.

Later on in his career, Carnap reiterated what he took to be Hume’s prime aim; namely, to “purify”, but not to “reject” the concept of causation, where the object of the purification was to free causation from the “component of necessity” (1974: 190). On Carnap’s view (1974: 201), Hume’s main contribution to the critique of causation is encapsulated in the thought that when one adds to a lawlike statement of the form “All *F*s are *G*s” the qualifier “and this holds with necessity”, one adds nothing of cognitive value since the alleged intrinsic necessity in any observed causal sequence cannot be observed, nor verified in any other way.

What emerges from all this is that, for empiricists, the concept of causation is intimately linked with the concept of law. And the latter

is intimately connected with the concept of regular (exceptionless) succession. Given that the reducing concept of regular succession is scientifically legitimate, the reduced concept of causation becomes legitimate too. But, as we have already seen on many occasions, regular succession (or correlation) does not imply causation. How could Schlick and Carnap have missed this point?

The following thought is available on their behalf. The operationalization of the concept of causation they were after was not merely an attempt to legitimize the concept of causation. Rather, it was part and parcel of their view that science aims at *prediction*. If prediction is what *really* matters, then the fact that there can be regularities, which are not causal in the ordinary sense of the word, appears to be irrelevant. A regularity can be used to predict a future occurrence of an event irrespective of whether it is deemed to be causal or not. The farmer can predict that dawn has broken on hearing the cock's crow irrespective of whether or not the crow causes the sunrise. In physics, one can *predict* the length of the pendulum's rod, given its period, irrespective of the causal connection between these two magnitudes. Correlations can serve prediction, even though they leave untouched some intuitive aspect of causation, according to which not all regularities are causal.

Schlick was very forthright about all this. Having expressed some concerns as to whether the concept of a law of nature can be properly explicated (cf. 1931, 1932), he noted that "the criterion of causality is successful prediction" (1932: 254, cf. also 1931: 184). In so far as a regularity can lead to successful predictions, Schlick insisted that it didn't matter whether this regularity satisfied any further criteria, which would deem it a law of nature (e.g. whether it was simple or complicated, or whether it made some explicit reference to particular times and places) (cf. 1931: 187). In a move reminiscent of the epistemic version of RVL (see section 5.4), Schlick noted that any regularity can be legitimately said to "behave as causality requires", provided that we are careful to note a single prediction issued by this regularity (1931: 188).

Carnap too noted that "causal relation means predictability" (1974: 192). But he was much more careful than Schlick in linking the notion of predictability – and hence, of causality – with the notion of the law of nature. For not all predictions are equally good. Some predictions rely on laws of nature, and hence they are

more reliable than others that rely on “accidental universals” (1974: 214). So, for Carnap, causation is not *just* predictability. It is more akin to subsumption under a universal regularity, that is, a law of nature. As he noted:

When someone asserts that A caused B, he is really saying that this is a particular instance of a general law that is universal with respect to space and time. It has been observed to hold for similar pairs of events, at other times and places, so it is assumed to hold for any time and place. (1974: 204)

Similarly, to say that event B was caused by event A is to say that “there are certain laws in nature from which event B can be logically deduced when they are combined with the full description of event A” (1974: 194). It seems reasonable to argue that, in contradistinction to Schlick, what Carnap was really after was the connection between causation and *explanation*. When we look for explanations, as opposed to predictions, we seem to look for something more than regularity, and relations of causal dependence might well be what we look for. So the thought suggests itself that what distinguishes between a causal regularity and a mere predictive one is their different roles in explanation. It appears, then, that the concept of explanation, and in particular of nomic explanation, can be the main tool for an empiricist account of causal dependence. As we shall see in the next section, this thought was made prominent by Hempel and Oppenheim’s deductive–nomological model of explanation.

8.2 Nomic expectability

What is it to explain a singular event *e*, for example, the explosion of a beer keg in the pub’s basement? The intuitive answer would be to provide the cause of this event: what brought about its occurrence. But is it enough just to cite another event *c*, for example, the rapid increase of temperature in the basement, in order to offer an adequate explanation of *e*? Explanation has to do with *understanding*. So an adequate explanation of event *e* (that is, of why *e* happened) should offer an adequate understanding of this happening. Just citing a cause would not offer an adequate understanding,

unless it was accompanied by the citation of a law that connects the two events. For just citing that c is the cause of e does not enhance our understanding of in virtue of what c occurred, nor of why e followed. Hence, it does not enhance our understanding of why e happened. Or, at least, so Hempel (1965) thought (cf. also Kim 1999: 11). According to Hempel the concept of explanation is primarily *epistemic*: to explain an event is to show how this event would have been *expected* to happen, had one taken into account the laws that govern its occurrence, as well as certain initial conditions. If one expects something to happen, then one is not surprised when it happens. Hence, an explanation amounts to the removal of the initial surprise that accompanied the occurrence of the event e . *Nomic expectability* is the slogan under which Hempel's account of explanation can be placed. In terms of the example stated above, the explosion of the beer keg is explained by citing the nomological connection between the rise of temperature and the increase of pressure.

Hempel systematized a long philosophical tradition by explicating the concept of explanation in terms of his deductive-nomological (DN) model. A singular event e (the *explanandum*) is explained if and only if a description of e is the conclusion of a valid deductive argument, whose premises, the *explanans*, involve essentially a lawlike statement L , and a set C of initial or antecedent conditions. The occurrence of the *explanandum* is thereby subsumed under a natural law. Schematically, to offer an explanation of an event e is to construct a valid deductive argument of the following form:

(DN)

Antecedent/initial conditions C_1, \dots, C_n

Lawlike statements L_1, \dots, L_n

Therefore, e event/fact to be explained (*explanandum*)

Before we examine this model in some detail, it is useful to locate it within the logical empiricist tradition we referred to in the previous section. At least since the mid 1930s, the logical empiricists thought that the prime aim of philosophy of science (indeed, of philosophy *simpliciter*) was to offer logical analysis of the syntax of basic scientific concepts. Clarification was thought to be effected by displaying the logical structure of these concepts. Philosophy was

indeed taken to amount to the *logic of science*. In a fashion similar to the use of logic in the clarification of problems in the foundations of mathematics, they took logic to be the essential tool for the analysis of scientific concepts. When it comes to the concept of explanation, they thought that a structural/syntactic account of this concept would ground its objectivity, since (a) it would display unambiguously its logical form, and (b) it would legitimize it, by showing how the concept of explanation could be meaningfully applied. Besides, as we saw in the previous section, they thought that the logical analysis of the structure of the concept of explanation would cast light on the *connection* between explanation and causation, where the latter was taken to be a slippery and problematic notion. The basic idea was that the concept of causation is linked with the concept of explanation via the laws that are needed to make an explanation cogent.

It is not hard to see how the Hempelian DN model is an integral part of this tradition.¹ A DN explanation is a special sort of a valid deductive argument – whose logical form is both transparent and objective – and conversely, the species of valid deductive arguments that can be DN explanations can be readily circumscribed, given only their form: the presence of lawlike statements in the premises is the characteristic that marks off an explanation from other deductive arguments. Hempel codified all this by offering 3 *plus* 1 conditions of adequacy for an explanation.

Conditions of adequacy:

1. The argument must be deductively valid.
2. The *explanans* must contain essentially a lawlike statement.
3. The *explanans* must have empirical content, i.e., they must be confirmable.
4. The *explanans* must be true.

The first three conditions are called “logical” by Hempel (1965: 247), because they pertain to the *form* of the explanation. The lawlike statement should occur essentially in the argument in the sense that the initial/antecedent conditions alone should not be enough for the derivation of the *explanandum*. On a first approximation at least, the characterization of a lawlike statement was taken to be purely syntactic (and hence, purely formal): a lawlike

statement is a universally quantified statement of the form “All *F*s are *G*s”. However, as we saw in detail in Chapter 5, this purely syntactic characterization suffers from insuperable problems, which made clear that, in the end, the analysis of the concept of explanation couldn’t be purely formal. In a certain sense, the third condition already suggests that the characterization of an explanation cannot be a purely logical or formal matter since the confirmability of the *explanans* (and especially of the lawlike statements that occur in them) depends on the kind of *predicates* that occur in them. Whether or not a statement is confirmable (i.e. whether or not it has empirical content) cannot just be a function of its logical form. Hempel was indeed aware of this, as he took pains to dismiss certain *teleological explanations* in biology in terms of purposive behaviour as not being “capable of empirical test” (1965: 256). But, initially, he paid little attention to the fact that an adequate account of explanation presupposes a theory of what kinds of predicate are confirmable. As we saw in section 5.4, this issue took centre stage after Goodman’s (1983) seminal work.

The final (fourth) condition of adequacy is “empirical”. Hempel rightly thought that it was a contingent matter whether the premises of an explanation were true or false. We can easily find ourselves in a situation in which an argument satisfies the first three conditions, and yet we may still wonder whether the premises are true. He called a DN argument that satisfies the first three conditions a “potential explanation”, that is, a valid argument such that, if it were also sound, it would explain the *explanandum*. He contrasted it with an “actual explanation”, which is a sound DN argument. So the fourth condition is what separates a potential from an actual explanation. The latter is the correct, or the true, explanation of an event. With the fourth condition, Hempel separated what he took to be the issue of “the logical structure of explanatory arguments” (1965: 249, note 3) from the empirical issue of what is the correct explanation of an event. But as we have already briefly noted, the structure of an explanatory argument cannot be purely logical. Indeed, if the issue of whether an argument was a potential explanation of an event was purely logical, it would be an a priori matter to decide that it was a potential explanation. But condition three shows that this cannot be a purely a priori decision. Without empirical information about the kinds of

predicates involved in a lawlike statement, we cannot decide whether the *explanans* have empirical content.

Sometimes the reference to laws in an explanation is elliptical and should be made explicit, or the relevant covering laws are too obvious to be stated. Hempel thought that a proper explanation of an event should use laws, and that unless it uses laws it is, in some sense, defective (perhaps in need of completion or in need of “further research” that will reveal the relevant laws (cf. 1965: 250)). It’s no accident that the DN model became known as “the covering law model” of explanation. In a slogan form, the main thesis could be that laws and only laws adequately explain the occurrence of singular events. Subsumption under laws is the hallmark of Hempelian explanation. So a lot turns on what exactly the laws of nature are and, as we have already seen in Chapters 5 to 7, this has proved to be a very sticky issue.

To highlight its relevance to the DN model, let me note that only *genuine* laws can explain. Accidents (although true) cannot. To use Hempel’s own example, suppose we want to explain why John Jones is bald. To this purpose, we can construct a DN argument whose *explanans* are the following two (true) statements: “John Jones is a member of the Greenbury School Board for 1964” and “All members of the Greenbury School Board for 1964 are bald” (1965: 339). Precisely because the major premise of the argument is true, but not lawlike, this argument lacks explanatory force. In contradistinction to this, the lawlike statement *All gases expand when heated under constant pressure* can adequately explain why a certain quantity of gas in a container expanded when it was heated under constant pressure. The relevant DN argument has explanatory force because the lawlike statement expresses a genuine law. But then without a robust distinction between laws and accidents, the DN model loses most of its putative force as a correct account of explanation.

8.3 The basic thesis

Hempel took his model to provide the correct account of *causal explanation*.² As he put it: “causal explanation is a special type of deductive nomological explanation” (1965: 300). Hempel (1965: 301) thought that not all DN explanations are causal. For instance,

low-level laws (e.g. Galileo's law) can be explained in a DN fashion, by showing that they logically follow from other higher-level laws (e.g. Newton's laws). But such explanations, although DN in character, are not causal. We shall deal with the issue of the explanation of laws in Chapter 10. For the time being, our focus will be what I shall call the *basic thesis* (BT):

(BT) All causal explanations of singular events can be captured by the deductive–nomological model.

So BT asserts: if Y is a causal explanation of a singular event, then Y is also a DN explanation of this event. BT might sound overly strong since Hempel did go on to offer non-deductive (i.e. statistical) accounts of explanation. But before readers make up their minds as to whether BT is indeed too strong as an interpretation of Hempel's view, they are advised to wait until Chapter 9.

The thought expressed by BT is firmly rooted in the empiricist project to legitimize – and demystify – the concept of causation by subsuming it under the concept of explanation, which in turn is modelled on relations of deductive entailment. So when the claim is made that event c causes event e (e.g. that the sugar cube dissolved *because* it was immersed in water), this claim should be understood as follows: there are relevant laws L_1, \dots, L_n in virtue of which the occurrence of the antecedent condition c is nomologically sufficient for the occurrence of the event e . "This relation between causal factors and effects," Hempel noted, "is reflected in our schema (DN): causal explanation is, at least implicitly, deductive-nomological" (1965: 349). This view has important implications for the concept of causation. In elaborating BT, Hempel (1965: 350) noted that when we say that event c caused event e , "the given causal statement must be taken to claim by implication that an appropriate law or set of laws holds by virtue of which $[c]$ causes $[e]$ ".³

We shall come to some of the implications of Hempel's views in subsequent sections, but now it's time to review some of the notorious problems that the DN model faces. We have already dealt with the most important one; namely, the nature of the laws of nature. By way of reminder, let us note that the best shot that empiricists can take on this issue is the MRL view. This goes a long way towards solving the problem of laws in a Humean way, but as

we have seen, it is not without problems of its own. In Chapter 10, when we see Friedman's (1974) model of explanation-as-unification, we shall highlight how the Hempelian DN model can be brought together with the MRL view of laws.

8.4 Enter causation

It has been a standard criticism of the DN model that, in so far as it aims to offer sufficient and necessary conditions for an argument to count as a *bona fide* explanation, it patently fails. For there are arguments that satisfy the structure of the DN model, and yet fail to be *bona fide* explanations of a certain singular event. Conversely, there are *bona fide* explanations that fail to instantiate the DN model. In what follows, we shall examine the relevant counter-examples and try to see how a Hempelian can escape, if at all, from them. To get a clear idea of what they try to show, let me state their intended moral in advance. This is that the DN model fails precisely because it leaves out of the explication of the concept of explanation important considerations about the role of *causation* in explanation. In other words, the moral of the counter-examples is that BT fails: there is more to the concept of causation than what can be captured by DN explanations. With this in mind, let us look at the counter-examples.

8.5 Of flagpoles and their shadows

The first class of counter-examples, which aims to show that the DN model is insufficient as an account of explanation, is summarized by the famous flagpole-and-shadow case. Suppose that we construct a DN explanation of why the shadow of a flagpole at noon has a certain length. Using the height of the pole as the initial condition, and employing the relevant nomological statements of geometrical optics (together with elementary trigonometry), we can construct a deductively valid argument with a statement of the length of the shadow as its conclusion. So we can DN-explain why the length of the shadow at noon is what it is. But as Sylvain Bromberger (1966) observed, we can reverse the order of explanation: we can explain the height of the flagpole, using the very same nomological statements, but (this time) with the length of the

shadow as the initial condition. Surely this is not a *bona fide* explanation of the height of the pole, although it satisfies the DN model.⁴ For it is not a *causal* explanation of the height of the pole: although the height of the pole is the *cause* of its shadow at noon, the shadow does not cause the flagpole to have the height it does.

This counter-example can be easily generalized by exploiting the functional character of some lawlike statements in science: in a functional law, we can calculate the values of each of the magnitudes involved in the equation that expresses the law by means of the others. Hence, given some initial values for the *known* magnitudes, we can calculate, and hence DN-explain, the value of the *unknown* magnitude. Suppose, for instance, that we want to explain the period T of a pendulum. This relates to its length l by the functional law: $T = 2\pi\sqrt{l/g}$. So we can construct a DN argument whose conclusion is some value of the period T and whose premises are the above law statement together with some value l of the length as our antecedent condition. Suppose, instead, that we wanted to explain the length of the pendulum. We could construct a DN argument similar to the above, with the length l as its conclusion, using the very same law statement but, this time, conjoined with a value of the period T as our antecedent condition. If, in the former case, it is straightforward to say that the length of the pendulum *causes* it to have a period of a certain value, in the latter case, it seems problematic to say that the period causes the pendulum to have the length it does.

Put in more abstract terms, the DN model allows explanation to be a symmetric relation between two statements; namely, the statement that expresses the cause and the statement that expresses the effect. So, given the relevant nomological statements, an effect can DN-explain the cause as well as conversely. If we take causation to be an asymmetric relation, then the DN model seems unable to capture fully the nature of causal explanation, despite Hempel's contentions to the contrary. The DN model allows the relation of explanatory dependence to go either from the cause to the effect or from the effect to the cause; but the relation of causal dependence has a definite direction. Causal explanation should surely respect this asymmetry. So the above counter-examples suggest that there is, after all, a conclusive argument against the empiricist thesis BT that causal explanation can be fully captured by arguments describing explanatory dependencies in a DN fashion.

The foregoing considerations have discredited the DN model. But there is still an interesting philosophical issue: if someone wanted to stick to the DN account of explanation and its concomitant claim to cover *all* causal explanation, if, that is, someone wanted to defend BT, what sort of moves would be available to them?

It should be stressed that the counter-examples we have seen so far do not contradict BT. They contradict the converse of BT, a thesis that might be called (+):

- (+) All deductive–nomological explanations of singular events are causal explanations.

But neither Hempel nor his followers endorse (+). He fully accepted the existence of non-causal DN explanations of singular events (cf. 1965: 353).⁵ The counter-examples do not dispute that causal explanation is a subset of DN explanation. What they claim is that the DN model licenses apparently inappropriate explanations, their inappropriateness being that they fail to be causal. This claim does *not* contradict BT. Still, the above counter-examples do show something important; namely, that unless causal considerations are imported into DN explanatory arguments, they fail to distinguish between legitimate (because causal) and illegitimate (because non-causal) explanations. So the task faced by the defender of the DN model is to show what could be added to a DN argument to issue in legitimate (causal) explanations. Schematically put, we should look for an extra *X* such that DN model + *X* = causal explanation. What could this *X* be?

8.5.1 Laws of coexistence and laws of succession

One move, made by Hempel (1965: 352) is to take *X* to be supplied by the law statements that feature in a DN explanation. To this end, Hempel relied on a distinction between *laws of coexistence* and *laws of succession*. A *law of coexistence* is the type of law in which an equation links two or more magnitudes by showing how their values are related to one another. Laws of coexistence are *synchronic*: they make no essential reference to time (i.e. to how a system or a state evolves over time); hence, they state how the

relevant magnitudes relate to each other at any given time. The law of the pendulum, Ohm's law and the laws of ideal gases are relevant examples. A *law of succession*, on the other hand, describes how the state of a physical system changes over time. Galileo's law and Newton's second law would be relevant examples. In general, laws of succession are described by differential equations. Given such an equation, and some initial conditions, one can calculate the values of a magnitude over time. Laws of co-existence display a kind of symmetry in the dependence of the magnitudes involved in them, but laws of succession do not. Or, at least, they are not symmetric given the fact that *earlier* values of the magnitude determine, via the law, *later* values.

Given this distinction, Hempel (1965: 352) argued that only laws of succession could be deemed causal. Laws of coexistence cannot. They do not display the time asymmetry characteristic of causal laws. But note now that the first type of counter-examples to the DN model, where there is explanatory symmetry but causal asymmetry, involves laws of coexistence. In such cases, the explanatory order can be reversed. But if these laws are *not* causal, then there is no problem: there is no causally relevant feature of these laws, which is not captured by relations of explanatory dependence. The length of the pendulum DN-explains its period no less than the period DN-explains the length. Given that the law of co-existence that connects the two magnitudes is *not* a causal law, we can easily say that none of the magnitudes *causes* the other: they just stand in some functional relation to each other. Hence, the thrust of the counter-example, namely, that by allowing the reversal of the explanatory order, the DN model misses out some important facts about the causal order, is neutralized.

That, then, is how Hempel replied to the asymmetry problem. The intuition may well be that the length of the pendulum *causes* its period, and not the other way around. But Hempel's retort is that the intuition is, to say the least, inconclusive. It appears that the period T of the pendulum is the "dependent variable" that is being controlled by changes of the length l – the "independent variable". But, appearances to the contrary, both the length and the period are just two functionally dependent variables: none of them is really independent of the other in the sense that it can be changed independently of the other. We cannot change the length of the

pendulum independently of changing its period, since if we want to change the length, we need first to *stop* the pendulum. And if we stop the pendulum, we change its period no less than we can change its length. So the extra *X* that should be added to a DN argument in order to ensure that it is a causal explanation has to do with the asymmetric character of some laws. Only asymmetric laws are causal, and can issue in causal explanations. So, DN explanation + asymmetric laws (of succession) = causal explanation.

8.5.2 Explanation and manipulation

Still, there seems to be something unsatisfactory in Hempel's reply, for the thought will be, we do make causal ascriptions, even when laws of coexistence are involved. It was, after all, the *compression* of the gas that caused its pressure to rise, even though pressure and volume are two functionally dependent variables related by a law of coexistence. This seems to be a valid objection. However, the following answer is available to someone who wants to remain Hempelian, attributable basically to von Wright (1973). Strictly speaking, when laws of coexistence are referred to in a DN explanatory argument, the explanation can be symmetric: we can explain the values of magnitude *A* by reference to the values of magnitude *B*, and conversely. But, Hempel's defender might go on, in particular *instances* of a DN explanatory argument with a law of coexistence, this symmetry can be (and is) broken. How the symmetry is broken – and hence how the direction of explanation is determined – depends on which of the functionally dependent variables is actually *manipulated*. Take, for instance Boyle's law, which says that, at constant temperature, the pressure of gas is inversely proportional to its volume. One can explain the increase of pressure by citing the decrease of volume, and conversely. When we want to say that, on a particular occasion, it was the decrease of pressure that *caused* the expansion of the gas, what grounds this causal claim is the fact that the factor that was manipulated was the pressure (and not the volume). On other occasions, when, for instance, we manipulate the volume, what caused the increase of pressure was the compression of the volume of the gas.

So, when laws of coexistence are involved, the symmetry that DN explanations display can be broken in different ways in order

to capture the actual causal order (i.e. what causes what on a particular occasion). An appeal to manipulability can also show how Bromberger-type counter-examples can be avoided. A DN model that cites the length of the shadow as the explanation of the height of the flagpole should not count as a *bona fide* explanation. For, although the length of the shadow is functionally dependent on the height of the pole, only the height of the pole is really manipulable. One can create shadows of any desirable length by manipulating the heights of flagpoles, but the converse is absurd.⁶ Manipulability can then be seen as the sought-after supplement X to the DN model which determines what the causal order is across different symmetric contexts in which a DN argument is employed. So DN explanation (with functional laws) + manipulability = causal explanation.

Yet, as we have already seen in section 3.4, the concept of manipulation is clearly causal. This means that advocates of DN explanation who summon von Wright's help can at best have a Pyrrhic victory. For they are forced to employ irreducible causal concepts in their attempt to show how a DN model of explanation can accommodate the intuitive asymmetry that explanatory arguments can possess.

Suppose that we granted that an appeal to manipulability breaks the symmetry of DN explanatory arguments with laws of coexistence. One could legitimately think that we could also construct DN explanations with *laws of succession* in which the actual causal order was reversed: we could DN-explain *earlier* values of a magnitude, no less than we could DN-explain *later* values. So, for instance, we can DN-explain why Mars will be in a certain position in the sky next week by stating Kepler's first law, and certain antecedent and initial conditions about Mars's orbit today, thereby capturing the correct causal order. But we can easily reverse the order of the explanation. We can DN-explain *past* positions of Mars in the sky by using the very same law and certain initial conditions about its *present* position. It would be absurd, however, to say that Mars's present position causes its having been in a certain place in the past. So it appears that even when it comes to laws of succession, the DN-model of explanation can leave out important facts about the causal order. It's not surprising that Hempel thought this objection fails (cf. 1965: 351–2). Laws of succession,

he thought, are causal in so far as (and because) they have a built-in *temporal* order: it is earlier values that are connected, via a law of succession, with later values. When the temporal order is reversed, we no longer have a causal explanation. But this does not mean that we have no explanation at all.

Be that as it may, the important issue is that in order for Hempel to avoid these counter-examples he needs (a) to build into causation a temporal order; (b) to deem causal only the laws that display, by default, a temporal direction; and (c) to insist that even when laws of succession are involved in a DN explanation, this explanation is not causal *unless* the initial conditions are temporally earlier than the *explanandum*. This is a controversial issue because it presupposes that there is a preferred temporal order (which has past events preceding future events) that is independent of the causal order (that is, of the direction of causation). According to this view, which can be readily traced back to Hume's account of causation, causation has a fixed temporal direction: causes precede in time their effects. So there cannot be *backward causation*, that is, causal relations in which the effect precedes in time the cause.

We cannot even start addressing this issue here in any proper sense. Many philosophers, for instance, think that the direction of causation cannot be settled a priori. Even if, in the actual world, the causal order has a preferred, forward-looking, direction, it is possible that in other possible worlds this direction could be reversed. In any case, the most important difficulty with building a preferred temporal order into causation is that any attempt at an explanation of the preferred direction of time should either be forfeited, or be cast in non-causal terms. In particular, some philosophers – most notably Hans Reichenbach (1956) – have tried to explain the temporal order (namely, the direction of time) in terms of the causal order (namely, the direction of causal relations). The relevance of all this to the attempt to rescue the DN account of causal explanation should be obvious: by building a preferred temporal order into the DN model, we cannot, on pain of circularity, offer a causal account of the temporal order. Yet it is not clear how fatal this problem is to the advocate of the DN model of causal explanation.⁷

Let me sum up. Counter-examples such as those considered in this section leave BT unscathed. Recall that BT says: if *Y* is a causal explanation of a singular event, then *Y* is also a DN explanation of

this event. If we accept that the counter-examples examined in this section succeed, they show what we may call the *insufficiency thesis* (IT):

- (IT) If Y is a deductive–nomological explanation of a singular event, then Y is not necessarily also a causal explanation of this event.

BT is consistent with IT. Yet IT does raise important collateral problems about the ability of the DN model to differentiate in non-causal terms between legitimate and illegitimate explanations of singular events.

8.6 Of ink-stains and hidden treasures

The popular philosophical claim that the DN model leaves important causal considerations out of the picture is also supported by a second class of counter-examples. These aim to show that satisfaction of the DN model is not a necessary condition for *bona fide* explanations. In fact, these counter-examples aim directly to discredit BT. Remember that BT says, in effect, that the concept of cause can operate legitimately only as a part of a suitable deductive-nomological argument. So saying that *c* causes *e* will be an elliptical claim, unless it is offered as an abbreviation for a full-blown DN argument. This view has been challenged by Scriven. As he (1958: 194) put it: “Producing a law is one way, not necessarily more conclusive, and usually less easy than other ways of supporting the causal statement.” He supported this point by the famous example of the explanation of the ink stain on the carpet. Citing the fact that the stain on the carpet was *caused* by inadvertently knocking over an ink bottle from the table, Scriven (1962: 90) argues, “is the explanation of the state of affairs in question, and there is no nonsense about it being in doubt because you cannot quote the laws that are involved, Newton’s and all the others”. So his point is that there can be fully legitimate *causal* explanations that are not DN in character. Instead, they are causal stories, that is, stories that give causally relevant information about how an effect was brought about, without referring to any laws, and without having the form of a deductive argument. Collaterally, it has been a standard

criticism of the Hempelian model that it wrongly makes all explanations arguments. A main criticism is that citing a causal mechanism can be a legitimate explanation of an event without having the form of a Hempelian DN argument (cf. Salmon 1989: 24).

One can accept Scriven's objection without abandoning either the DN model of explanation; or BT. The fact that the relevant nomological connections may not be fully expressible in a way that engenders a proper deductive explanation of the *explanandum* merely shows that, on some occasions, we shall have to make do with what Hempel called "explanation sketches" instead of full explanations. Explanation-sketches can well be ordinary causal stories that, as they stand, constitute incomplete explanations of an event *E*. But these stories can, nonetheless, be completed by taking account of the relevant laws that govern the occurrence of the event *E*.⁸ Scriven's point, however, seems to be more pressing. It is that a causal explanation can be *complete*, without referring to laws (cf. 1962: 94). So he directly challenges Hempel's assumption that all causal explanation *has to be* nomological. Scriven insists that explanation is related to understanding and that the latter might, but won't necessarily, involve reference to laws. So he proposes (1962: 95): "a causal explanation of an event of type [*E*], in circumstances [*R*] is exemplified by claims of the following type: there is a comprehensible cause [*C*] of [*E*] and it is understood that [*C*]s can cause [*E*]s". But, a Hempelian might argue, it is precisely when we move to the nomological connection between Cs and Es that we understand how Cs can cause Es.⁹

One important implication of the DN model is that *there is no genuine singular causal explanation* (cf. Hempel 1965: 350, 361–2). Scriven's own objection can be taken to resonate with the singularist approach to causation (see Chapter 2). It might be taken to imply that a singular causal explanation of an event-token (e.g. the staining of the carpet by ink) is a complete and fully adequate explanation of its occurrence. Since the DN model denies that there can be legitimate singular *causal* explanations of events, what is really at stake is whether causal stories that are not nomological can offer legitimate explanation of singular events. So what is at stake is BT.

Note that there is an ambiguity in the singularist approach. What does it mean to say that there is *no* nomological connection between two event-tokens *c* and *e* that are nonetheless such that *c* causally explains *e*? It might mean one of the following two things:

(a) there are no relevant event-types under which event-tokens *c* and *e* fall such that they are nomologically connected to each other; (b) even if there is a relevant law, we don't (can't) know it; nor do we have to state it explicitly in order to claim that the occurrence of event-token *c* causally explains the occurrence of event-token *e*.

The first option is vulnerable to the following objection. One reason why we are interested in identifying causal facts of the form *c* causes *e* (e.g. heating a gas at constant pressure causes its expansion) is that we can then *manipulate* event-type *C* in order to bring about the event-type *E*. But the possibility of manipulation requires that *there is* a nomological connection between types *C* and *E*. It is this nomological connection that makes possible bringing about the effect *e* by manipulating its causes. Hence, if causation is to have any bite, it had better instantiate laws.¹⁰ So, I think, the singularist's assertion should be interpreted to mean the second claim above, namely, that even if there is a law connecting event-types *C* and *E*, we don't know it; nor do we have to state it explicitly in order to claim that the occurrence of event-token *c* causally explains the occurrence of event-token *e*. Given this understanding, it might seem possible to reconcile the singularist approach with a Humean one. As we saw in section 2.6, this is precisely the line taken by Davidson (1967). On his view, all causation is nomological, *but* stating the law explicitly is not required for causal explanation.

Considering this idea, Hempel noted that when the law is not explicitly offered in a causal explanation, the statement "*c* causes *e*" is incomplete. In making such a statement, one is at least committed to the view that "*there are* certain further unspecified background conditions whose explicit mention in the given statement would yield a truly general law connecting the 'cause' and the 'effect' in question" (1965: 348). But this purely existential claim does not amount to much. For, as Hempel went on to say, the foregoing claim is comparable to having "a note saying that there is a treasure hidden somewhere" (1965: 349). Such a note would be useless unless "the location of the treasure is more narrowly circumscribed". So the alleged reconciliation of the singularist approach with the Humean will not work, unless there is an attempt to make the covering law explicit. But this will inevitably take us back to forging a close link between stating causal dependencies and stating actual regularities.¹¹

To sum up: if the counter-example considered in this section were correct, it would establish the thesis that the DN model is not necessary for causal explanation. Let's call this thesis UNT. UNT says: if *Y* is a causal explanation of a singular event, then *Y* is not necessarily a DN explanation of this event. UNT, if true, would contradict BT. But we haven't yet found good reasons to accept UNT.

8.7 What if things had been different?

This is as good a place as any to mention a very recent attempt to challenge UNT by Woodward (1997, 2000). In section 7.2, we examined in some detail Woodward's view that laws of nature are best seen as expressing invariant relations among magnitudes. As the reader might recall, a central point made by Woodward was that invariant relations outrun lawlike relations. So there are invariant relations (especially in the special sciences) that are not laws. When it comes to explanation, Woodward's main idea is that invariant relations can be important to explanation, even if they do not express laws. Given that laws are, on Woodward's view, a species of invariant relations, the DN model can be explanatory. So his objection to the DN model is precisely that it *requires* laws, where just invariant relations will do. As he put it, "nonlawful explanation is possible" (1997: S29). Why are invariant relations important to explanation? According to Woodward, invariant relations tell us not just what happens, but also what would have happened if certain interventions were made. So they can answer a network of "what-if-things-had-been-different questions", thereby placing the *explanandum* within a pattern of counterfactual dependencies. For instance, the law of ideal gases is said to be explanatory not because it renders a certain *explanandum* (e.g. that the pressure of a certain gas increased) nomically expected, but because it can tell us how the pressure of the gas would have changed, had the antecedent conditions been different. The explanation proceeds by locating the *explanandum* "within a space of alternative possibilities" (2000: 209). As noted above, Woodward argues that a basic advantage of his approach is that it shows how there can be explanation in the special sciences, where invariant relations abound, but exceptionless regularities are scarce.

Woodward's account deserves much more attention than I can give it now. It certainly seems to open up the possibility of a genuine challenge to the DN model, while respecting (and grounding) the intuition that laws *are* explanatory. I will leave it to the reader to examine it further. The only worry I want to register is something related to what was pointed out in section 7.2.1: if invariance-under-interventions is not robust enough to distinguish between laws and accidents, then it seems that the invariance of a relation may not be enough to ground its role in explanation.

8.8 Explanation and prediction

There is an interesting consequence of the DN model that Hempel, at least, welcomed. A DN explanation of a singular event amounts to a *prediction* of its occurrence. It is easy to see that the *explanans* of a DN argument are sufficient to entail that the *explanandum* will occur, and hence to *predict* its occurrence. As Hempel (1965: 249) noted: "It may be said, therefore, that an explanation of a particular event is not fully adequate unless its explanans, if taken account of in time, could have served as a basis for predicting the event in question." In fact, he went as far as to state that "whatever will be said . . . concerning the logical characterisation of explanation or prediction will be applicable to either, even if only one of them should be mentioned" (*ibid.*). This important implication of the DN model has become known as the *explanation/prediction-symmetry thesis*. Hempel welcomed this implication precisely because he thought it could distinguish between good explanations, which have this potential predictive force, and bad or inadequate explanations (e.g. in prescientific discourse), which lack this force. Yet it has been observed that the alleged symmetry between prediction and explanation breaks all too often in good explanations. There can be explanations that do not have predictive force and, conversely, there can be perfectly legitimate predictions that offer no explanation. So in so far as the DN model makes the symmetry between explanation and prediction a necessary and sufficient condition for a good explanation, it fails.

A standard counter-example to the view that all predictions can also be explanations is the following. Well-functioning barometers can be used to predict an upcoming storm. Yet neither

barometers on their own, nor the hypothesis that they correlate well with storms, explain why the storm has occurred. It is a drop in the atmospheric pressure that explains the occurrence of the storm. In fact, both the drop of the barometer and the subsequent storm are common effects of the same cause; namely, the fall in the pressure. So there can be predictions without explanation. Once more, what went amiss here was the actual causal order. The prediction is not an explanation precisely because the predictive hypothesis does not identify the right cause for the occurrence of the predicted event. The counter-examples that aim to show that there can be explanations without predictions are far more interesting because they relate to the objection about the laws of succession mentioned in section 8.5.2. A standard type of counter-example is this. Suppose that we DN-explain a past event (e.g. the position of Mars two months ago) by using Kepler's first law and the present position of the planet as the initial condition. Although this is an explanation licensed by the DN model, it does not amount to a prediction of Mars's position, since predictions are forward-looking. The relevant DN explanation of Mars's position two months ago may retrodict this position, but it does not predict it. Once again, we can see that the failure of the explanation/prediction symmetry thesis is tied to the alleged failure of the DN model to align the explanatory order with the causal order.¹²

8.9 Causal histories

In Lewis's very important work (1986b), he takes causal explanation of a singular event to consist in providing some information about its causal history. In most typical cases, it is hard to say of an effect *e* that its cause was *the* event *c*. Lots of things contribute to bringing about a certain effect. Take the case that Lewis discusses. What caused a certain car accident? Well, a number of factors contributed: the driver was tipsy, it was raining, the visibility was poor, the road was slippery, the corner was blind, and more. Each of them contributed to the crash. They jointly caused the crash. And the story does not stop here. For each of these causes had its own causes, and so forth. So, Lewis says, all these factors comprise the *causal history* of the effect. This history is a huge causal net in which the effect is located. So, to explain why this event happened,

we need to offer some information about this causal net. This is “explanatory information” (1986b: 25). A *full* explanation consists in offering the whole causal net. But this full explanation is hardly ever possible. Nor, Lewis thinks, is it necessary. Often some chunk of the net will be enough to offer an adequate causal explanation of why a certain singular event took place.

This account merits more attention than I can offer it now. But there is a feature of it that connects with the prospects of Hempel’s DN model, and especially with BT. So we shall focus on this. Lewis (1986b: 221–4) thinks that there is no such thing as non-causal explanation of singular events. That is, he endorses the following thesis:

(CE) All explanation of singular events is causal explanation.

Recall that BT says:

All causal explanation of singular events can be captured by the deductive–nomological model.

If we added BT to CE, then it would follow that

(CE*) All explanation of singular events can be captured by the deductive–nomological model.

Hempel and others accept BT but disagree with CE. But let’s leave that to one side. The question that concerns us now is whether Lewisians could accept BT, and hence whether they could also accept CE*. Lewis criticizes the DN model of explanation on the basis that some DN explanations fail to capture relevant causal connections and hence fail to offer a genuine causal explanation. To this extent, Lewis is right in saying that “we do not, at least not yet, have a DN analysis of causation” (1986b: 234). But, does Lewis’s account of causal explanation violate BT? Or is his view of causal-explanation-as-information-about-causal-histories compatible with BT? Lewis (1986b: 235–6) asks: “is it . . . true that any causal history can be characterised completely by means of the information that can be built into DN arguments?” Obviously, if the answer is positive, then BT is safe. Lewis expresses some scepticism about a

fully positive answer to the above question. He thinks that if his theory of causation, based on the notion of counterfactual dependence (see section 3.3), is right, then there can be genuinely singular causation. Yet he stresses that in light of the fact that the actual world seems to be governed by a “powerful system of (strict or probabilistic) laws, . . . the whole of a causal history could in principle be mapped by means of *DN*-arguments . . . of the explanatory sort” (1986b: 236). He adds:

if explanatory information is information about causal histories, as I say it is, then one way to provide it is by means of *DN* arguments. Moreover, under the hypothesis just advanced [i.e. the hypothesis that the actual world is governed by a powerful system of laws], there is no explanatory information that could not in principle be provided in that way. To that extent the covering-law model is dead right. (*Ibid.*)

So, BT is safe for a Lewisian, at least if it is considered as a thesis about causal explanation in the actual world. What, then, is Lewis’s disagreement with the DN model? There is a point of principle and a point of detail (or so I think). The point of principle is this. BT has not been discredited. But, if I understand Lewis correctly, he thinks that it has been wounded. It may well be the case that if *Y* is a causal explanation of a singular event, then *Y* is also a DN explanation of this event. Lewis does not deny this (cf. 1986b: 239–40). But, in light of the first set of counter-examples, and the concomitant IT – see end of section 8.5.2 – BT might have to be modified to BT’:

(BT’) All causal explanation of singular events can be captured by suitable instances of the deductive–nomological model.

The modification is important. For it may well be the case that what instances of the DN model are suitable to capture causal explanations might well be specifiable only “by means of explicitly causal constraints” (1986b: 236). And if this is so, then the empiricists’ aspiration to capture causal concepts by the supposedly unproblematic explanatory concepts seems seriously impaired.

The point of detail is this. Take BT’ to be unproblematic. It is still the case, Lewis argues, that the DN model has wrongly searched for

a “unit of explanation” (1986b: 238). But there is no such unit: “It’s not that explanations are things we may or may not have one of; rather, explanation is something we may have more or less of” (*ibid.*). So, although Lewis agrees that a full DN explanation of an individual event’s causal history is both possible and most complete, he argues that this ideal is chimerical. It is the “ideal serving of explanatory information” (1986b: 236). But, “other shapes and sizes of partial servings may be very much better – and perhaps also better within our reach” (1986b: 238). This is something that the advocate of the DN model need *not* deny. What is really at stake is not the point of detail, but the point of principle.

But it’s now time to look at Hempel’s attempt to generalize his model of explanation to cover cases of *explanandum*, which have only some probability to happen.

9 Statistical explanation

Hempel's pioneering work on explanation consists really in his analysis of the circumstances under which we can explain events whose occurrence is not certain (cf. 1965: 376–412). Hempel's models of statistical explanation were really the first systematic treatment of the subject. In this chapter, we shall examine these models, and a major alternative to them. Our focus, in the end, will be the implication of statistical explanation for the Humean approach to causation. For there is a firm thought that there are causal relations between event-types that are not linked by strict (deterministic) laws. For instance, we do believe that smoking causes lung cancer, even if there is no strict law that says that whoever smokes develops lung cancer. Lots of delicate issues need to be dealt with here but, in line with the apology offered in the Introduction, I will leave most of them untouched.

9.1 Explaining statistical regularities

Suppose that we want to explain a statistical regularity; namely, the fact that in a large collection of atoms of the radioactive isotope of Carbon-14 (C^{14}) approximately three-quarters of them will very probably decay within 11,460 years. This, Hempel (1965: 380–81) observed, can be explained *deductively* in the sense that its description can be the conclusion of a valid deductive argument, whose premises include a statistical nomological statement. The general claim above follows deductively from the statement that every C^{14} atom has a probability of 0.5 of disintegrating within any period of

5,730 years (provided that it is not exposed to external radiation). There is no big mystery here. A valid deductive argument can have as its conclusion a statistical generalization provided that one of the premises also contains some suitable probabilistic statement. Hempel called this account the *deductive–statistical* (DS) model of explanation. Salmon (1989: 53) rightly observes that the DS model is just a species of the DN model, when the latter is applied to the explanation of statistical regularities. So most of the problems that the DN model faced are inherited by the DS model.

But there is more to statistical explanation than the DS model can cover. For, as Hempel (1965: 381) noted, we are also interested in explaining *singular events* whose probability of happening is less than unity. Suppose, to exploit one of his own examples (cf. 1965: 382), that Jones has suffered from septic sore throat, which is an acute infection caused by bacteria known as *streptococcus hemolyticus*. He takes penicillin and recovers. There is no strict (deterministic) law, which says that whoever is infected by streptococcus and takes penicillin will recover quickly. Hence, we cannot apply the DN model to account for Jones's recovery. Nor can we apply the DS model, since what we want to explain is an individual event, not a statistical regularity. How are we to proceed?

9.2 Explanation of likely events

Suppose, Hempel says, that there is a statistical generalization of the following form: whoever is infected by streptococcus and takes penicillin has a very high probability of recovery. Let's express this as follows:

$\text{prob}(R/P\&S)$ is very high,

where R stands for quick recovery, P stands for taking penicillin and S stands for being infected by streptococcus germs. We can then say that given this statistical generalization, and given that Jones was infected by streptococcus and took penicillin, the probability of Jones's quick recovery was high. So, Hempel thought, we have *inductive* grounds to expect that Jones will recover. We can then construct an inductive argument that constitutes the basis of the explanation of an event whose occurrence is governed by a statisti-

cal generalization. This is then the birth of Hempel's *inductive-statistical* (IS) model. Let a stand for Jones, and let R , P and S be as above. Applied to Jones's case, the IS explanations can be stated thus:

$$\begin{array}{ll}
 (1) & \\
 Sa \text{ and } Pa & \\
 \hline \hline
 \text{prob}(R/P\&S) \text{ is very high} & \text{[makes practically certain} \\
 Ra & \text{(very likely)]}
 \end{array}$$

More generally, the logical form of an IS explanation is this:

$$\begin{array}{ll}
 (\text{IS}) & \\
 Fa & \\
 \hline \hline
 \text{prob}(G/F) = r, \text{ where } r \text{ is high (close to 1)} & [r] \\
 Ga &
 \end{array}$$

The double line before the conclusion indicates that it is an *inductive* argument. The conclusion follows from the premises with high probability. The strength r of the inductive support that the premises lend to the conclusion is indicated in square brackets. Being an inductive argument, an IS explanation is such that its premises may be true and its conclusion false. That is, if we have as premises Fa and $\text{prob}(G/F) \neq 1$, we cannot deduce Ga , no matter how high $\text{prob}(G/F)$ might be. In fact, Scriven seized upon this possibility to argue that "statistical statements are too weak – they abandon the hold on the individual case. . . . An event can rattle around inside a network of statistical laws" (quoted by Hempel 1965: 391, n. 14). Hempel rightly replied that the fact that an IS explanation rests on an inductive argument does not imply that its premises cannot explain the conclusion. After all, Ga did occur and we can explain this by saying that, given the premises, we would have expected Ga to occur. Besides, holding fast to the view that the only legitimate explanation must be non-statistical makes a mockery of scientific practice, which is fraught with statistical explanations.

The IS model inherits a number of important features of the DN model. The IS model makes explanations arguments, albeit inductive. It also understands explanation as nomic expectability. To

explain an event is still to show how this event would have been *expected* (with high probability) to happen, had one taken into account the statistical laws that govern its occurrence, as well as certain initial conditions. The IS model needs an essential occurrence of law statements in the *explanans*, albeit expressing statistical laws.

A natural objection might be that probabilistic-statistical laws can explain the characteristics of large samples, but cannot explain anything about an individual case. So suppose that we flip a fair coin 10,000 times. The law that the probability of a fair coin's landing heads is 0.5, together with the assumption that each tossing of the coin is statistically independent from any other, can be used to explain why the number of heads in the 10,000 tossings is somewhere between 4,900 and 5,100. According to the laws of statistics, this outcome has a high probability of 0.95. But can this explain why a *specific* individual tossing of the coin landed heads? Hempel didn't claim that it could. For, after all, the probability of each individual tossing of the coin landing heads is 0.5, which is not high. But he did claim that if the probability of the occurrence of an individual event is *very high*, then its occurrence *could* be explained no less than the general fact we saw above. This is made vivid when we consider an example such as this. Suppose that we have an urn with 1,000 balls, 999 of which are white and one black. Suppose we draw a ball *a*, (*Da*), and it is white, (*Wa*). The probability $\text{prob}(W/D)$ of a ball being white given that it is drawn (with replacement) from the urn is 0.999. Then, the following IS explanation of the drawn ball's *Da* being *Wa* seems perfectly natural.

$$\frac{Da}{\frac{\text{prob}(Wa/Da) = 0.999}{Wa}} \quad [0.999]$$

Hempel's requirement of high probability is essential to his IS model. It's this requirement that makes the IS model resemble the DN model, and it is also this requirement that underwrites the idea that an IS explanation is a good inductive argument. Yet this requirement is actually one of the major problems that the IS explanation faces. For it seems clear that we also need to explain events whose occurrence is *not* probable, but which, however, do occur. In the example mentioned above, it seems that we would also need to

explain the drawing of the unique black ball, even if its drawing was very unlikely, its probability being only 0.001. Or, although only one in one hundred babies get born with a very low birthweight (i.e. less than 1,500 grams), the fact that a certain baby was born with weight less than 1,500 grams is still in need of explanation. Richard Jeffrey (1969) highlighted this weakness of the IS model by noting that the requirement of high probability is not a necessary condition for statistical explanation. We must look elsewhere for the hallmark of good statistical explanation. In particular, if the requirement of high probability is relaxed, then statistical explanations are no longer arguments.

Is the requirement of high probability sufficient for a good statistical explanation? The answer is also negative. To see why, we should look at some aspects of the statistical regularities that feature in the IS model. Suppose, to use one of Salmon's (1989: 58) examples, we explain why Jones recovered from a common cold within a week by saying that he took a large quantity of vitamin C. We can then rely on a statistical law, which says that the probability of recovery from common colds within a week, given taking vitamin C, is very high. The formal conditions for an IS explanation are met and yet the argument offered is not a good explanation of Jones's recovery from a common cold, for the statistical law is no good. It is *irrelevant* to the explanation of recovery since common colds, typically, clear up after a week, irrespective of the administration of vitamin C. This suggests that more stringent requirements should be in place if a statistical generalization is to be explanatory. High probability is not enough.

It is noteworthy that the specific example brings to light a problem of IS that seems to be detrimental. The reason why we think that the foregoing statistical generalization is not explanatory is that we, rightly, think that it fails to capture a *causal connection* between recovery from common colds and the administering of vitamin C. That two magnitudes (or variables) are connected with a high-probability statistical generalization does not imply that they are connected causally. Even when the connection is not statistical but deterministic, it still does not follow that they are causally connected. Correlation does not imply causation. To say the least, two magnitudes (or variables) might be connected with a high-probability statistical generalization (or by a deterministic one)

because they are effects of a common cause. So the causal arrow does not run from one to the other, but instead from a common cause to both of them.

It might be thought that the IS model is not aimed at causal explanation. Indeed, Hempel refrained from explicitly connecting IS explanation with causal explanation (cf. 1965: sections 3.2, 3.3 and pp. 58–9, 67–9). However, in Hempel (1965: 393), he toyed with the idea that the IS model offers “a statistical-probabilistic concept of ‘because’ in contradistinction to a strictly deterministic one, which would correspond to deductive-nomological explanation”.¹ So it’s fair to say that in so far as the IS model aims to capture a sense of statistical (or probabilistic) causation, it fails.

For simplicity, I shall disregard the issue of when a statistical generalization expresses a causal law. I will make only a few scattered remarks, when it seems absolutely necessary. One cannot even begin to address this issue properly, unless one is prepared to spend a lot of time and use a lot of space. This task – although within the scope of the present book – should be deferred to a different occasion. In fact, although we devoted Chapters 5 to 7 to a detailed investigation of when a deterministic regularity is a law of nature, I will say no more about the relevant issue when it comes to statistical generalizations.

But I cannot resist making a very general point about the use of the concept of probability in the IS model. This concept enters twice in the IS model. It is used to describe the nature of an IS argument, and it is used in the characterization of statistical generalizations. These are, for Hempel at least, distinct interpretations of the concept of probability. The first is *logical*, while the second is *physical*. The interested reader should look at Carnap’s monumental (1950) work for the details of this distinction. But on a good approximation, the difference is this. When it is said that the conclusion (*explanandum*) of an IS argument follows from the premises (*explanans*) with high probability, the concept of probability refers to the logical relation between the premises and the conclusion. It is, that is, a logical relation among statements. This is not a relation of deductive entailment, but it is a relation of inductive support or confirmation. Carnap (1950) did think that this notion can be usefully seen as *partial entailment*. But Hempel (1965: 385) thought that the more neutral notion of “(degree of) inductive support” is enough. When,

however, we say that a generalization states a probabilistic relation among some variables, for example, that $\text{prob}(G/F) = r$, the concept of probability has a different interpretation. It states, roughly, that *in the long run*, the proportion of those F s that are G s, among all F s, is approximately r . This is called a *frequentist* interpretation of probability, since it ties probabilities to (limiting) relative frequencies. The generalization does not merely state that the proportion of F s that has been observed to be G is r . Rather its meaning can be best seen if we envisage a kind of random experiment capable of repetition, whose G is a possible result. For instance, we can take F to be a coin being repeatedly tossed and G to be its coming up heads. Then the claim that $\text{prob}(G/F) = r$ (where $r = 0.5$ in the case at hand) says that if we kept repeating F (that is, if we kept tossing the coin), then, *in the long run*, it is practically certain that the relative frequency of G (coming up heads) among the outcomes will be very close to r ($= 0.5$). All this has a precise mathematical definition that cannot be given here. But it is easy to see that this account of probability is objective. Understood this way, the statistical generalization states an objective fact about the relative frequency of an attribute in a population.

Difficulties with this objective understanding of probability have led many philosophers to take another route in an attempt to characterize probabilities objectively. This is to think of probabilities as *objective chances*. On this view, to say that $\text{prob}(G/F)$ is not to say something about relative frequencies, but instead something about the objective chance of F being G . A big difference between these two objective approaches is that, on the chance account, it makes perfect sense to talk about single-case probabilities, that is, of the probability of an event happening, even if this event happens just once. On the frequentist account, it simply does not make any sense to speak of probabilities where frequencies cannot be specified. These issues have important implications about the nature of statistical generalizations as well as the nature of statistical explanation. But, as I said, I won't pursue them further.²

9.2.1 Adding new information

Enough has been said so far to bring to light the grave difficulties of the IS model. But there is another one, which will pave the way for

a better understanding of the nature of statistical explanation, and its relation to causation. Hempel (1965: 394) called this problem “the ambiguity of inductive-statistical explanation”.

Valid deductive arguments have the property of *monotonicity*. If the conclusion Q follows deductively from a set of premises P , then it will also follow if further premises P^* are added to P . Inductive arguments, no matter how strong they may be, lack this property: they are *non-monotonic*. The addition of extra premises P^* to an inductive argument may even remove the support that the original set of premises P conferred on the conclusion Q . In fact, the addition of extra premises P^* to an inductive argument may be such that the *negation* of the original conclusion becomes probable. Take our stock example of Jones’s recovery from streptococcal infection and refer to its IS explanation (1), in section 9.2. Suppose, now, that Jones was, in fact, infected by a germ of streptococcus that was resistant to penicillin. Then, Jones’s taking penicillin cannot explain his recovery. Actually what is now likely to happen is that Jones won’t recover from the infection, despite the fact that he took penicillin, and despite the fact that it is a true statistical generalization that most people who take penicillin recover from streptococcus infection. The addition of the extra premise that Jones was infected by a penicillin-resistant strain (Ta) will make it likely that Jones won’t recover ($not-Ra$). For now the probability $\text{prob}(not-R/P\&S\&T)$ of non-recovery ($not-R$) given penicillin (P), streptococcal infection (S), and a penicillin-resistant germ (T) is very high. So:

(2)

Sa and Pa

Ta

$\text{prob}(not-R/P\&S\&T)$ is close to 1

$not-Ra$

[makes practically
certain (very likely)]

The non-monotonic nature of IS explanation makes all this possible. (1) and (2) are two arguments with mutually consistent premises and yet incompatible conclusions. It is this phenomenon that Hempel called the “ambiguity” of IS explanation. What is ambiguous is in what reference class we should include the *explanandum*. Given that it may belong to lots of difference reference classes, which one shall we choose? In which reference class shall

we include Jones's illness? Will it be, for instance, the reference class of people who took penicillin, or the reference class of people who took penicillin and were infected by a resistant germ of streptococcus, or the class of those who took penicillin, were infected by a resistant germ of streptococcus and are over 80 years old, and so on? The problem is precisely that different specifications of the reference class in which the *explanandum* might be put will lead to different estimations of the probability of its occurrence. Consider the following: what is the probability that an individual lives to be 80 years old? The answer will vary according to which reference class we place the individual. To name but a few reference classes that confer different probabilities on this individual event: the class of those who smoke 40 cigarettes a day for 40 years; the class of those who smoke 40 a day for 40 years but live in a very healthy environment and exercise regularly; the class of those who don't smoke at all but have a very weak heart, and so on.

The problem we are discussing is accentuated if we take into account the fact that, even if there was an objectively correct reference class to which an individual event belongs, in most realistic cases when we need to explain an individual event, we won't be able to know whether the correct identification of the reference class has been made. We will place the individual event in a reference class in order to IS-explain its occurrence. But will this be the *right* reference class, and how can we know of it? This is what Hempel (1965: 395) called the *epistemic version* of the ambiguity problem. The result of this is that an IS explanation should always be relativized to a body *K* of currently accepted (presumed to be true) beliefs.

Note that the problem of ambiguity does not arise in the case of the DN explanation, which we considered in Chapter 8. The premises of a DN argument are *maximally specific*. If it is the case that *All Fs are Gs*, then no further specification of *Fs* will change the fact that they are *Gs*. All *Fs* and *Hs* are still *Gs*, as still are all *Fs* and *Hs* and *Ls*. If all humans are mortal, then all humans over 40 are mortal, and all humans over 40 who exercise are mortal, and so on. So, if all *Fs* are *Gs*, then every individual *a* that is *F*, will also be *G*, irrespective of what other properties it might have (or of what other reference classes it might belong to). As the jargon goes, if all *Fs* are *Gs*, no further partition of the reference class *F* can change the probability of an instance of *F* to be also an instance of *G*, this

probability being already equal to unity. On the contrary, in an IS explanation, further partitions of the reference class F can change the probability that an instance of F is also an instance of G .

This suggests that we may introduce the Requirement of Maximal Specificity (RMS) to IS explanation. Roughly, to say that the premises of an IS explanation are maximally specific is to say that the reference class to which the *explanandum* is located should be the narrowest one. In order to see how we should specify this in a better way, let's go back to Jones's recovery. Suppose that Jones was, in fact, infected by a penicillin-resistant germ of streptococcus (T). Suppose also that $\text{prob}(R/P\&S) = r$. This premise is not maximally specific because it leaves out the fact T . The reference class of $P\&S$ (i.e. the class of those infected by streptococcus and who took penicillin) is not the narrowest one. It can be further partitioned into two classes: $P\&S\&T$ (i.e. the class of those infected by streptococcus and who took penicillin and were infected by a penicillin-resistant strain of streptococcus) and $P\&S\&\text{not-}T$ (i.e. the class of those infected by streptococcus and who took penicillin and were *not* infected by a penicillin-resistant strain of streptococcus). Suppose that $\text{prob}(R/P\&S\&T) = r_1$. This narrower class is *relevant* to the IS explanation of Jones's illness because $\text{prob}(R/P\&S\&T)$ is different from $\text{prob}(R/P\&S)$, that is, because $r \neq r_1$. If it was not relevant, then $\text{prob}(R/P\&S\&T)$ would be equal to $\text{prob}(R/P\&S)$. Intuitively, if the factor T were not relevant to Jones's illness, then its addition to factors R and S would *not* change the probability of Jones's recovery. This can lead to a formal characterization of RMS.

Suppose that the set P of premises of an IS explanation of an individual event Fa imply that $\text{prob}(G/F) = r$. The set of premises P is maximally specific if, given that background knowledge K tells us that a also belongs to a subclass F_1 of F , and given that $\text{prob}(G/F_1) = r_1$, then $r = r_1$.³

Let's call a reference class *homogeneous* if it cannot be further partitioned into subclasses that violate RMS. Clearly, there are two concepts of homogeneity. The first is objective: there is no partition of the reference class into subclasses which violate RMS.⁴ The second is epistemic: we don't (currently) know of any partition that violates RMS. Hempel's version of RMS was the latter. Hence, IS explanation is always relativized to a certain body of background knowledge K , which asserts what partitions of the reference classes

are *known* to be relevant to an IS explanation of an individual event. The fact that IS explanations are always epistemically relative has made many philosophers think that the IS model cannot be an adequate model of statistical explanation (cf. Salmon 1989: 68ff). Alberto Coffa (1974: 69) was perhaps the first to point out that the IS model cannot tell us when a statistical explanation is true. But then the IS model functions as “a placebo which can only calm the intellectual anxieties of the unconscious user”. What we would need of a statistical explanation is an identification of the relevant features of the world that are nomically connected (even in a statistical sense) with the *explanandum*. The IS model is far from doing that, as the problem with RMS makes vivid.

The friends of statistical explanation face a dilemma. They might take the view that all genuine explanation is DN and hence treat statistical explanation as *incomplete* explanation. If, indeed, all explanation is DN, then the problem of the reference class (and of RMS) does not even arise. On this view, an IS explanation is a placeholder for a full DN explanation of an individual event. The statistical generalizations are taken to express our *ignorance* of how to specify the correct reference class in which we should place the *explanandum*. This approach is natural, if one is committed to determinism. According to determinism, every event that occurs has a fully determinate and sufficient set of antecedent causes. Given this set of causes, its probability of happening is unity. If we knew this full set of causes of the *explanandum*, we could use this information to objectively fix its reference class and we would, thereby, establish a true universal generalization under which the *explanandum* falls. If, for instance, the full set of causes of event-type *E* was the conjunction of event-types *F*, *G* and *H*, we could simply say that “All *F*s & *G*s & *H*s are *E*s”. So, on the view presently discussed, statistical generalizations simply express our ignorance of the full set of causes of an event. They are by no means useless, but they are not the genuine article. In section 11.2 we shall see how this view is elaborated by Kitcher (1989).

Alternatively, the friends of statistical explanation could take the view that there is *genuine* statistical explanation, which is nonetheless captured by a model different to the IS model. In order to avoid the pitfalls of the IS model, they would have to admit that there is a fact of the matter as to the *objectively homogeneous* reference class

in which a certain *explanandum* belongs. But this is not enough for genuine statistical explanation, since, as we saw in the previous paragraph, the existence of an objectively homogeneous reference class is compatible with the presence of a universal law. So the friends of genuine statistical explanation should also accept that even within an objectively homogeneous reference class, the probability of an individual event's occurring is not unity. So they have to accept indeterminism: there are no further facts that, were they taken into account, would make this probability equal to unity. An example (cf. Salmon 1989: 76) will illustrate what is at issue here. Take a collection of radioactive C^{14} atoms whose half-life is 5,730 years. This class is as close to being objectively homogeneous as it can be. No further partitions of this class can make a subclass of C^{14} atoms have a different half-life time. But what is important here is that the law that governs the decay of C^{14} atoms is *indeterministic*. The explanations that it licenses are genuinely statistical, because the probability that an atom of C^{14} will decay within 5,730 years is irreducibly 0.5. In genuine statistical explanation, there is no room to ask certain why-questions. Why did this *specific* C^{14} atom decay? If indeterminism is true, there is simply no answer to this question.

The issue that crops up then is the following. Can the friends of genuine statistical explanation offer an adequate model of it? Let's devote the next section to this issue.

9.3 Making a difference

Take an event-type E whose probability of happening given the presence of a factor C (i.e. $\text{prob}(E/C)$) is r . In judging whether a further factor C_1 is relevant to the explanation of an individual event that falls under type E , we look at how taking C_1 into account affects the probability of E happening. If $\text{prob}(E/C \ \& \ C_1)$ is different from $\text{prob}(E/C)$, then the factor C_1 is *relevant* to the occurrence of E . Hence, it should be *relevant* to the explanation of the occurrence of an individual event that is E . Let's say that:

- C_1 is positively relevant to E , if $\text{prob}(E/C \ \& \ C_1) > \text{prob}(E/C)$;
- C_1 is negatively relevant to E , if $\text{prob}(E/C \ \& \ C_1) < \text{prob}(E/C)$;
- and C_1 is irrelevant to E , if $\text{prob}(E/C \ \& \ C_1) = \text{prob}(E/C)$.

Judgements such as the above seem to capture the intuitive idea of *causal relevance*. We rightly think, for instance, that the colour of one's eyes is causally irrelevant to one's recovery from streptococcus infection. We would expect that one's probability of recovery (R) given streptococcus infection (S) and penicillin (P), that is, $\text{prob}(R/P \& S)$, will be unaffected, if we take into account the colour of one's eyes (B). So, $\text{prob}(R/P \& S) = \text{prob}(R/P \& S \& B)$. Analogously, we would think that the fact that one is infected by a penicillin-resistant strain of streptococcus (T) is causally relevant to one's recovery (in particular, its lack). So we would expect that $\text{prob}(R/P \& S \& T) < \text{prob}(R/P \& S)$. These thoughts, together with the fact that the requirement of high probability is neither necessary nor sufficient for a good statistical explanation, led Salmon (1984, Salmon *et al.* 1971) to suggest a different conception of statistical explanation. The main idea is that we explain the occurrence of an individual event by citing certain statistical-relevance relations. In particular,

a factor C explains the occurrence of an event E , if
 $\text{prob}(E/C) > \text{prob}(E)$ – which is equivalent to
 $\text{prob}(E/C) > \text{prob}(E/\text{not-}C)$.

This came to be known as the *statistical-relevance* (SR) model.⁵ Where an IS explanation involves just one probability value, the SR model suggests that explanation compares two probability values. As the jargon goes, we need to compare a *posterior probability* $\text{prob}(E/C)$ with a *prior probability* $\text{prob}(E)$. Note that the actual values of these probabilities do not matter. Nor is it required that the posterior probability be high. All that is required is that there is a *difference*, no matter how small, between the posterior probability and the prior. Suppose, for example, that the prior probability $\text{prob}(R)$ of recovery from streptococcus infection is quite low, say 0.001. Suppose also that when one takes penicillin, the probability of recovery $\text{prob}(R/P)$ is increased by only 10 per cent. So, $\text{prob}(R/P) = 0.01$. We would not, on the IS model, be entitled to explain Jones's recovery on the basis of the fact that he took penicillin. Yet, on the SR model, Jones's taking penicillin is an explanatory factor of his recovery, since $\text{prob}(R/P) > \text{prob}(R)$. (Equivalently, $\text{prob}(R/P) > \text{prob}(R/\text{not-}P)$.)

Let's not go into the (very important and somewhat technical) details of the SR model. The interested reader is advised to look at Salmon's seminal work (1984: 36–47, 1989: 62–7). I shall only note one important feature of it, which paves the way to the entrance of *causation* into statistical explanation. Suppose that taking penicillin is explanatorily relevant to quick recovery from streptococcus infection. That is, $\text{prob}(R/P) > \text{prob}(R)$. Can we then, without further ado, say that taking penicillin causes recovery from streptococcus infection? Not really. For one might be infected by a penicillin-resistant strain (T), thus rendering one's taking penicillin totally ineffective as a cure. So, if we take T into account, it is now the case that $\text{prob}(R/P \& T) = \text{prob}(R/T)$. The further fact of infection by a penicillin-resistant germ renders *irrelevant* the fact that penicillin was administered. The probability of recovery given penicillin and infection by a penicillin-resistant germ is equal to the probability of recovery given infection by a penicillin-resistant germ. When a situation like this occurs, we say that factor T *screens off* R from P .

This relation of screening off is very important. Take the example discussed in the previous chapter. There is a perfect correlation between well-functioning barometers (B) and upcoming storms (S). The probability $\text{prob}(S/B)$ that a storm is coming up given a drop in the barometer is higher than the probability $\text{prob}(S)$ that a storm is coming up. So, $\text{prob}(S/B) > \text{prob}(S)$. It is in virtue of this relationship that barometers can be used to predict storms. Can we then, using the SR model, say that the drop of the barometer explains the storm? Worse, can we say that it causes the storm? No, because the correlation between a drop of the barometer and the storm is *screened off* by the fall of the atmospheric pressure. Let's call this A . It can be easily seen that $\text{prob}(S/B \& A) = \text{prob}(S/A)$. The presence of the barometer is rendered irrelevant to the storm, if we take the drop of the atmospheric pressure into account. Instead of establishing a causal relation between B and S , the fact that $\text{prob}(S/B) > \text{prob}(S)$ points to the further fact that the correlation between B and S exists because of a *common cause*. It is typical of common causes that they screen off the probabilistic relation between their effects. But a factor can screen off a correlation between two others, even if it's not their common cause. Such was the case of infection by a penicillin-resistant germ discussed above.

So if the probabilistic relations endorsed by the SR model are to establish genuine explanatory relations among some factors C and E , it's not enough to be the case that $\text{prob}(E/C) > \text{prob}(E)$. It is also required that this relation is not screened off by further factors. Put more formally:

C explains E if (i) $\text{prob}(E/C) > \text{prob}(E)$ [equivalently, $\text{prob}(E/C) > \text{prob}(E/\text{not-}C)$]; and (ii) there are no further factors H such that H screens off E from C , i.e., such that $\text{prob}(E/C\&H) = \text{prob}(E/H)$.

The moral of all this is that relations of statistical relevance do not imply the existence of causal relations. The converse seems also true, as the literature on the so-called *Simpson paradox* makes vivid. But we shall not go into this.⁶ Correlations that can be screened off are called *spurious*.

There should be no doubt that the SR model is a definite improvement over the IS model. Of course, if we go for the SR model, we should abandon the dominant empiricist idea that explanations are arguments. We should also question the claim that statistical generalizations are really necessary for statistical explanation. For an SR explanation is not an argument. Nor does it require citing statistical laws. Rather, as Salmon (1984: 45) put it, it is "an assembly of facts statistically relevant to the explanandum, regardless of the degree of probability that results". Besides, the SR model makes clear how statistical explanation can be seen as a species of causal explanation. For if the relevant SR relations are to be explanatory, they have to capture the right causal dependencies between the *explanandum* and the *explanans*. But it also paves the way for the view *that there is more to causation than relations of statistical dependence*. Salmon himself has moved from the claim that all there is to statistical explanation can be captured by specifying relations of statistical relevance to the claim that, even if we have all of them, we would still need to know something else in order to have genuine explanation; namely, facts about "causal relationships" (1984: 45). His latest view is this:

the statistical relationships specified in the S-R model constitute the *statistical basis* for a bone fide scientific explanation,

but . . . this basis must be supplemented by certain *causal factors* in order to constitute a satisfactory scientific explanation. (1984: 34)

So, according to Salmon (1984: 22), relations of statistical relevance must be explained by causal relations, and not the other way around.⁷ As we saw in section 4.2, his favoured account of causal relations is given in terms of unveiling the causal mechanisms, be they deterministic or stochastic, that connect the cause with its effect.

9.4 Explanation of unlikely events

Do deductivism and indeterminism mix? Can, that is, one think that although all explanation is, in essence, deductive, there is still space to explain essentially chance events? Railton's (1978, 1981) *deductive-nomological-probabilistic* (DNP) model of probabilistic explanation is a very important attempt to show how this can happen. So we must not fail to take a good look at it.

Being dissatisfied with the epistemic ambiguity of the IS model, and accepting the view that there should be space for the explanation of unlikely events, Railton (1981: 160) suggested that a legitimate explanation of a chance *explanandum* should consist in

- (a) a "law-based demonstration that the explanandum had a particular probability of obtaining"; and
- (b) a claim that, "by chance, it did obtain".

Take the case of a very unlikely event such as a Uranium-238 nucleus u decaying to produce an alpha-particle. The mean-life of a U^{238} nucleus is 6.5×10^9 years, which means that the probability p that such a nucleus will produce an alpha-particle is vanishingly small. Yet events like this *do* happen, and need to be explained. Railton (1981: 162–3) suggests that we construct the following two-step explanation of its occurrence.

The *first* step is a straightforward DN explanation of the fact that nucleus u has a probability p to alpha-decay during a certain time-interval Δt .

- (1a) All U^{238} nuclei not subjected to external radiation have probability p of emitting an alpha-particle during any time-interval Δt .
- (1b) u was a U^{238} nucleus at time t and was not subjected to any external radiation during time-interval $[t, t + \Delta t]$.

Therefore

- (1c) u has a probability p to alpha-decay during time-interval $[t, t + \Delta t]$.

Now, this step does not yet explain why the particular nucleus u alpha-decayed. It only states the probability of its decay. So, Railton says, the *second* step is to add a “parenthetic addendum” (1981: 163) to the above argument. This addendum, which is put *after* the conclusion (1c), says:

- (1d) u did indeed alpha-decay during the time-interval $[t, t + \Delta t]$.

If, in addition, the law expressed in premise (1a) is explained (derived) from the underlying theory (quantum mechanics, in this example), then, Railton (1981: 163) says, we have “a full probabilistic explanation of u ’s alpha-decay”. This is an instance of a DNP explanation.

The addendum (1d) is not an extra premise of the argument. If it were, then the explanation of why nucleus u alpha-decayed would be trivial. So the addendum has to be placed *after* the conclusion (1c). Still, isn’t there a feeling of dissatisfaction? Have we really explained why u did alpha-decay? If we feel dissatisfied, Railton says, it will be because we are committed to determinism. If, on the other hand, we take indeterminism seriously, there is no further fact of the matter as to why nucleus u alpha-decayed. This is a genuine chance event. Hence, nothing else could be added to steps (1a)–(1d) above to make them more explanatory than they already are. Note that I have refrained from calling steps (1a)–(1d) an argument because they are not. Better, (1a)–(1c) is a deductively valid argument, but its conclusion (1c) is *not* the *explanandum*. The *explanandum* is the “addendum” (1d). But this does not logically follow from (1a)–(1c). Indeed, Railton defends the view that

explanations are not necessarily arguments. Although arguments (and in particular DN arguments) “play a central role” in explanation, they “do not tell the whole story” (1981: 164). The *general schema* to which a DNP explanation of a chance event (G_{e,t_0}) conforms is this (cf. 1981: 163):

- (2a) For all x and for all t ($F_{x,t} \rightarrow \text{Probability } (G)_{x,t} = r$)
 (2b) A theoretical derivation of the above probabilistic law
 (2c) F_{e,t_0}

 (2d) $\text{Probability } (G)_{e,t_0} = r$
 (2e) (G_{e,t_0}).

(2e) is the “parenthetic addendum”, which is *not* a logical consequence of (2a)–(2d). As for (2a), Railton stresses that the probabilistic generalization must be a genuine law of nature. The explanation is true if both the premises (2a)–(2c) *and* the addendum (2e) are true.

There are a number of important features of the DNP model that need to be stressed.

- It shows how the DN model is a limiting case of the DNP model. In the case of a DN explanation, (2e) is just the conclusion of the DN argument – so it is no longer a “parenthetic addendum”.
- It shows that all events, no matter how likely or unlikely they may be, can be explained in essentially the same way. In schema (2) above, the value of probability r is irrelevant. It can be anywhere within the interval $(0,1]$. That is, it can be anything other than zero.
- It shows that single-case probabilities, such as the ones involved in (2a), can be explanatory. No matter what else we might think of probabilities, there are cases, such as the one discussed in Railton’s example, in which probabilities can be best understood as fully objective chances.
- It shows how probabilistic explanation can be fully objective. Since (2a)–(2d) is a valid deductive argument, and since the probability involved in (2a) is “a law-full, physical single-case” probability (cf. 1981: 166), the DNP account does not fall prey to the objections that plagued the IS model. There is no

problem of ambiguity, or epistemic relativization. Single-case probabilities need no reference classes, and by stating a law, premise (2a) is maximally specific.⁸

- It shows how probabilistic explanation can be freed of the requirement of nomic expectability as well as of the requirement that the *explanandum* had to occur. So, it accommodates genuinely chance *explananda*.
- By inserting premise (2b), it shows in an improved way how explanation can be linked with understanding.

Since this last point is of some special importance, let us cast some more light on it.

9.4.1 An explanatory web

The Hempelian tradition took explanation to be the prime vehicle for understanding in science. But, as we saw in Chapter 8, it restricted the understanding of why an *explanandum* happened to showing how it should have been expected to happen, given the relevant laws. In particular, it demanded that understanding should proceed via the construction of arguments, be they DN, DS or IS. Railton's DNP model suggests that understanding of why an *explanandum* happened cannot just consist in producing arguments that show how this event had to be expected. The occurrence of a certain event, be it likely or not, is explained by placing this event within a *web* of:

inter-connected series of law-based accounts of all the nodes and links in the causal network culminating in the explanandum, complete with a fully detailed description of the causal mechanisms involved and theoretical derivations of all covering laws involved. (1981: 174)

In particular, explanation proceeds also with elucidating the *mechanisms* that bring about the *explanandum*, where this elucidation can only be effected if we take into account the relevant theories and models. Railton (1981: 169) rightly protested against the empiricist view that all this extra stuff, which cannot be captured within a rigorous DN argument, is simply “*marginalia*, incidental

to the ‘real explanation’, the law-based inference to the explanandum”. He does not doubt that an appeal to laws and the construction of arguments are important, even indispensable, features of explanation. But he does doubt that they exhaust the nature of explanation.

The explanatory web mentioned in the previous paragraph is what Railton (1981: 167) aptly calls “an ideal DNP text”. Schema (2) (see previous section) is just a condensed summary of the ideal DNP text. If premise (2b) was to be fully unpacked, then it would comprise the full ideal text relevant to the *explanandum*. Of course, such ideal texts cannot be written. But this is irrelevant. For, as Railton (1981: 174) notes, what we (better, the scientists) need for explanation and understanding is not the full text. *Full* understanding of why an event occurs would require the full ideal DNP text. But this is more of a regulative ideal than what, in practice, we need and should strive for. In practice, what we (or the scientists) need and should strive for is “explanatory information” relevant to the *explanandum*. Such information, if indeed it is information relevant to the *explanandum*, will be part of the ideal DNP text. By producing such parts, no matter how underdeveloped and incomplete they may be, scientists understand why a certain *explanandum* happens. Finding more and more bits of the ideal texts, we move closer to the ideal of a full understanding. The fact that we never reach the latter, a fact which is not unlikely to be the case, does not imply that we have not gained some genuine understanding by the explanatory information we already have.

The idea of an ideal DNP text is extremely fruitful. But there seems to be a problem with it. There is no way to judge a priori what an ideal DNP text would look like. In particular, there is no a priori reason to think that an ideal text would contain information other than what is captured by nomological connections (be they deterministic or probabilistic) among several event-types. So it may be the case that the ideal DNP text is a nomological text. As we shall see in Chapter 11, this ideal text may well be the text that the *web-of-laws* approach to laws (see section 5.6) would entitle us to write, were we ever able to write it. So it may be a bit premature to say that the ideal text will contain explanatory information not exclusively couched in terms of law-based arguments, as it is equally premature to say that the ideal text will contain information about causal

mechanisms, which could not be captured by relevant nomological connections and DN explanatory arguments. So Humeans can be happy with the notion of an ideal text, since it does not necessarily seem to commit them to connections of explanatory dependence that they wouldn't accept.

It might be said that Railton's idea of an ideal text stems from an idealization of current scientific practice. In current practice, we see scientists explaining not just (or mainly) by producing DN arguments, or, more broadly by appealing to laws. They explain also by citing causal mechanisms, or by producing stories that do not make essential reference to laws. This may well be so. But this talk might be elliptic. So it might be that the content of the ideal text could be mapped by means of law-based arguments. To be sure, Railton (1981: 176) does say that according to his account of explanatory texts, "their backbone is a series of law-based deductions". He goes as far as to call his "the nomothetic account of scientific explanation" (1981: 176). However, he notes: "it is . . . difficult to dispute the claim that many proffered explanations succeed in doing some genuine explaining *without* either using laws explicitly or (somehow) tacitly asserting their existence" (*ibid.*).

9.4.2 Contrastive explanation

Before we move on, it is important to note that Railton's view that there can be explanations of chance events can be backed up by Lewis's (1986b: 229–31) account of "contrastive explanation". Suppose we ask the straightforward why-question: why *p*? Why, for instance, did this aircraft crash? As we saw in section 8.9, Lewis (1986b) suggests that this *plain* why-question is answered by offering causal information about the particular crash. But occasionally we ask *contrastive* why-questions: why *p*, rather than *q*? Why, for instance, did the joyrider take Jones's car rather than Smith's? When we ask contrastive questions, we (implicitly) contrast two causal histories. One is the actual causal history of the actual *explanandum* (the stealing of Jones's car), while the other is the unactualized (counterfactual) causal history of the unactualized event (the stealing of Smith's car). So, in explaining why *p* rather than *q* happened, we need to appeal to factors in *p*'s actual causal history that are *not* also factors in *q*'s (unactualized) causal history.

We need to do this because we look for a difference in the two causal histories that made p rather than q happen. So if we find out that both Jones and Smith have automatic cars, this factor, being present in both causal histories would not be relevant to the explanation of why it was Jones's car that was stolen rather than Smith's. But if we find out that Jones left his car unlocked in the night but Smith didn't, then this is a relevant difference. We can say that what caused the theft of Jones's car rather than Smith's is that Jones left it unlocked in the night.⁹

All this is useful in dismissing certain requests for explanation of chance events. Lewis agrees with Railton that certain plain questions about chance events (e.g. why did this U^{238} nucleus alpha-decay?) can be fully answered by offering DNP arguments. But his theory of contrastive explanation can also show why we should look no further for an explanation of why a chance event happened. A further explanation of why a chance event happened would be a call for an answer to a contrastive why-question: why did this U^{238} nucleus alpha-decay *rather than* not? This kind of contrastive why-question, Lewis (1986b: 230–31) says, is ill posed, when it comes to genuine chance events. For the alpha-decaying of a U^{238} nucleus is a genuine chance effect: there is no relevant difference in the actual causal history of its decay and the counterfactual causal history of its remaining stable. The two causal histories are identical (in fact they can both be captured by the very same DNP story) up to the time of the decay; yet according to the one the U^{238} nucleus did alpha-decay and according to the other it didn't. If indeterminism is true, then this may be all there is to say about the occurrence of a chance event (but also see section 11.2 for some more discussion of this issue).

This chapter has just scratched the surface of a cluster of important issues concerning the nature of statistical explanation. But no more was promised. So, we can now move back to deductivism and see how laws can be explained. This is the task of Chapter 10.

10 Explanation of laws

10.1 Explanatory ascent

Necessitarians (see Chapter 6) insist that Humeans cannot adequately show how laws explain their instances. Armstrong, for instance, notes that when we explain why all observed *F*s have been *G*s by stating that *All F*s are *G*s, we “explain something by appealing to a state of affairs part of which is the thing to be explained” (1983: 40). “But”, he adds, “a fact cannot be used to explain itself”. His point is that a generalization *A*: “All *F*s are *G*s” is equivalent to the conjunction of the following two statements: *A*₁: “All *observed F*s are *G*s” and *A*₂: “All *unobserved F*s are *G*s”. So, when we try to explain *A*₁ by reference to *A*, we presuppose *A*₁ and hence we cannot explain it. Necessitarians solve this problem by what Earman (1984: 215) has called an “ontological ascent”: the further fact that *F*-ness implies *G*-ness explains why all *F*s are *G*s as well as why all observed *F*s have been *G*s.

Innocuous though it may sound, this point is very important. Where necessitarians inflate their ontology to explain why regularities hold, Humeans must find a different explanation. They should engage in what Earman (*ibid.*) called “ascent of explanatory level”. Ultimately, it should be further regularities that explain lower-level regularities and their observed patterns. If Humeans succeed in doing that, they can avoid the unwanted to them ontological ascent. But can they? The intuitive idea is clear and forceful: explain why low-level laws hold by reference to more fundamental laws. But the implementation of this idea has proved to be much more demanding than it first seems.

There are two sets of problems here. First, we need to explain the circumstances under which a law can be said to explain another law. Secondly, we need to distinguish between more fundamental and less fundamental laws. Both of these problems lead us naturally to the attempts made by the descendants of the Hempelian DN model of explanation to show how laws can explain laws. In this chapter we shall take another look at the idea of a deductive systematization of the knowledge of the world. In particular, we shall see that the idea of deductive systematization can bring together the MRL approach to laws and the account of explanation offered by Friedman (1974) and Kitcher (1981, 1989). Briefly put, Friedman and Kitcher have tried to offer two things: first, a way to characterize the nature of explanatory dependence; secondly, a way to characterize the nature of the best deductive systematization. Both of these are brought together under a central concept: *unification*. As regards the nature of explanatory dependence, it is still derivation; but, not any kind of derivation. It is derivation within a *maximally unified theoretical system*. As regards the nature of the best deductive systematization, it too is the systematization that *maximally unifies a theoretical system*. Explanation and best deductive systematization become one.

10.2 Explanation and unification

When Hempel presented his DN model, he encountered the following difficulty (cf. 1965: 273). Suppose one wants to explain a low-level law L_1 in a DN-fashion. One can achieve this by simply subsuming L_1 under the more comprehensive regularity $L_1 \& L_2$, where L_2 may be any other law one likes. So, for instance, one can DN-explain Boyle's law by deriving it from the conjunction of Boyle's law with the law of adiabatic change. Although such a construction would meet all the requirements of the DN model, it wouldn't count as an explanation of Boyle's law. Saying that the conjunction $L_1 \& L_2$ is not more fundamental than L_1 would not help, for the issue at stake is precisely what makes a law more *fundamental* than another one. Intuitively, it is clear that the laws of the kinetic theory of gases are more fundamental than the laws of ideal gases. But if what makes them more fundamental *just* is that the latter are derived from the former, then the conjunction $L_1 \& L_2$

would also count as more fundamental than its components. Hempel admitted that he did not know how to deal with this difficulty. But this difficulty is very central to his project. The counter-example trivializes the idea that laws can be DN-explained by being deduced from other laws. Arguably, scientific explanation is centrally concerned with explaining regularities – perhaps more centrally than with explaining particular facts. Hence, the empiricist project should have to deal with the *problem of conjunction*.

10.2.1 Reducing the number of brute regularities

An intuitive idea is that a law is more fundamental than others, if it *unifies* them. But how exactly is *unification* to be understood? Why, that is, does $L_1 \& L_2$ not unify L_1 and L_2 ? Friedman (1974) was the first philosopher to address this problem systematically. According to him, explanation is closely linked with understanding. Now, *understanding* is a slippery notion. Intuitively, it relates to knowing the causes: how the phenomena are brought about. Yet for Humeans *understanding* should be analysed in non-causal terms. Friedman revived a long-standing empiricist tradition where *understanding* is linked to conceptual economy.¹ The basic thought is that a phenomenon is understood if it is made to fit within a coherent whole, which is constituted by some basic principles. If a number of seemingly independent regularities are shown to be subsumable under a more comprehensive law, then, the thought is, our understanding of nature is promoted, for the number of regularities that have to be assumed as *brute* is minimized. Some regularities, the fundamental ones, should still be accepted as brute. But the smaller the number of regularities that are accepted as brute, and the larger the number of regularities subsumed under them, the more we comprehend the workings of nature: not just what regularities there are, but also why they are and how they are linked to each other. After noting that in important cases of scientific explanation (e.g. the explanation of the laws of ideal gases by the kinetic theory of gases) “we have reduced a multiplicity of unexplained, independent phenomena to one”, Friedman added:

I claim that this is the crucial property of scientific theories we are looking for; this is the essence of scientific explanation –

science increases our understanding of the world by reducing the total number of independent phenomena that we have to accept as ultimate or given. A world with fewer independent phenomena is, other things equal, more comprehensible than with more. (1974: 15)

So explanation proceeds via unification into a compact theoretical scheme. The basic *unifiers* are the most fundamental laws of nature.² The explanatory relation is still deductive entailment, but the hope is that, suitably supplemented with the idea of minimizing the number of independently acceptable regularities, it will be able to deal with the conjunction problem. It may be a bit ironic that, as Friedman notes, a philosopher who expressed the core of Friedman's idea before was William Kneale. He stated that the:

explanation of laws by showing that they follow from other laws is a simplification of what we have to accept because it reduces the number of untransparent necessitations we need to assume. . . . What we can achieve . . . is a reduction of the number of independent laws we need to assume for a complete description of nature. (1949: 91–2)

The irony, of course, is that Kneale proposed this view within his account of laws as expressing *necessitating* principles (see section 6.1). There is a sense then in which Friedman's account of explanation is *neutral* with respect to the issue of the nature of laws. A necessitarian – like Kneale or even Armstrong – can accept that explanation is unification as well as that laws can be expressed within a unified deductive system. What is then at stake is *the order of dependence*. Where a Humean would say that some regularities count as laws because they (better, descriptions of them) are part of the best deductive system, a necessitarian would say that it is merely a symptom of laws that they get organized (or are organizable) in a deductive system. Yet it is also fair to say that Friedman's thought fits best to the empiricist-Humean project precisely because (a) it shows how the ontological ascent to universals *can* be avoided, and (b) it shows how explanatory relations can be mapped onto relations of deductive entailment.

10.2.2 Atomic sentences

In broad outline, Friedman's approach is the following. A lawlike sentence L_1 is acceptable independently of lawlike sentence L_2 , if there are sufficient grounds for accepting L_1 , which are not sufficient grounds for accepting L_2 . This notion of *sufficient grounds* is not entirely fixed. Friedman (1974: 16) states two conditions that it should satisfy:

- (a) If L_1 implies L_2 , then L_1 is not acceptable independently of L_2 .
- (b) If L_1 is acceptable independently of L_2 , and L_3 implies L_2 , then L_1 is acceptable independently of L_3 .

So the basic idea is that lawlike sentence L_1 is not acceptable independently of its logical consequences, but it is acceptable independently of other statements logically independent from it. So, for instance, given that $L_1 \& L_2$ entails L_1 , grounds for accepting $L_1 \& L_2$ are sufficient for accepting L_1 (and L_2). But the converse does not hold: grounds for accepting L_1 (or L_2) are not sufficient for accepting the conjunction $L_1 \& L_2$. This is not very illuminating, as Friedman admits. But a further step shows how this idea can be put to work in solving *the problem of conjunction*. Take a lawlike sentence L . Let us call a *partition* of L a set of sentences L_1, \dots, L_n such that

- (a) their conjunction is logically equivalent to L ; and
- (b) each member L_i of the set is acceptable independently of L .

Let us call *conjunctive* a sentence L that satisfies Friedman's two conditions, and, following Friedman, let us call "atomic" a sentence L that *violates* them. Given this, a lawlike sentence L explains lawlike sentences L_1, \dots, L_n , if L is "atomic". Conversely, a lawlike sentence L *fails* to explain lawlike sentences L_1, \dots, L_n , if L is *conjunctive*. We can now see how Friedman's account bars the mere conjunction $L_1 \& L_2$ of Boyle's law (L_1) with the law of adiabatic change (L_2) from explaining Boyle's law: the conjunction of the two laws is not an atomic sentence; it is a *conjunctive* sentence. It is partitioned into a (logically equivalent) set of independently acceptable sentences; namely, L_1 and L_2 . Conversely, we can see why Newton's law of gravity offers a genuine explanation, via unification, of Galileo's law,

Kepler's laws, the laws of the tides, etc. On Friedman's account, what is different between Newton's law and the mere conjunction L_1 & L_2 is that the content of Newton's law *cannot* be partitioned into a (logically equivalent) set of independently acceptable laws: the sentence that expresses Newton's law is "atomic".

Friedman is aware of the fact that what regularities are accepted independently of each other may vary with time. Hence, he relativizes his notions to a body K of accepted regularities – or, better, to a set K of sentences that describe the accepted regularities. So a sentence L is K -atomic if L is *not* partitioned into a set of independently acceptable sentences, which feature in K . If we take the set K of accepted lawlike sentences to be the ideal deductive system of the world, then we can see how Friedman's idea can be linked to the MRL view of laws. The most fundamental laws will be the axioms of the most unified system, that is, of the system that minimizes the number of independently acceptable regularities by the strongest and simplest set of atomic sentences.

All this sounds very promising. For now we seem to have a criterion by which we can stop *mere conjunctions* of already accepted regularities from counting as explanatory. Besides, we seem to have a neat account of unification: a lawlike sentence unifies a set of other lawlike sentences if it is atomic relative to this set. So Humeans seem able to deliver their promise: explanation proceeds via unification into a deductive system, whose axioms form the smallest and strongest set of independently acceptable atomic sentences, that is, a set from which (descriptions of) all other accepted regularities follow. Not only are the laws expressed by this set of axioms the fundamental laws of nature, but also whatever regularities cannot be derived from this set, without compromising its simplicity, are mere accidents. The bad news, however, is that Friedman's proposal will not deliver the goods.

10.2.3 Independent acceptability, properties and predictions

Kitcher (1976) has shown that Friedman's account does *not* offer a necessary condition for the explanation-as-unification thesis. His general point is that if, ultimately, explanation of laws amounts to derivation of lawlike sentences from other lawlike sentences, then in mathematical physics at least there will be many such derivations

that utilize more than one lawlike statement as premises. Hence, ultimately, there are *conjunctions* that are partitioned into independently acceptable lawlike statements which, nonetheless, explain other lawlike statements.

On the face of it, however, atomicity does offer a *sufficient* condition for genuine unifying, and hence explanatory, power. But if there were no atomic sentences, then this condition would be void. So can there be atomic lawlike sentences? At a purely syntactic level, there cannot be. For *any* sentence of the form “All *Fs* are *Gs*” can be partitioned into a logically equivalent set of sentences such as {“All (*Fs* & *Hs*) are *Gs*” and “All (*Fs* & *not-Hs*) are *Gs*”}. So the predicate *is a planet* can be partitioned into a set of logically equivalent predicates: *is a planet and is between the earth and the sun* (*F* & *H*) and *is a planet and is not between the earth and the sun* (*F* & *not-H*). Take, then, the statement that expresses Kepler’s first law; namely, that all planets move in ellipses. It follows that this can be partitioned into two statements: “All *inferior planets* move in ellipses” and “All *superior planets* move in ellipses”. A perfectly legitimate lawlike statement is partitioned into two other lawlike statements. Is it then “atomic”? Syntactic considerations alone suggest that it is not.

The natural retort for Friedman is to claim that although a partition of a lawlike statement is always possible at the syntactic level, it does not necessarily follow that there are no atomic statements, since the partition might not consist of independently acceptable statements. Take another example: Newton’s law of gravity can be partitioned into a set of three laws, one applying the inverse square law to two large bodies, another applying it to two small bodies, and a third applying it to one large and one small body (cf. Salmon 1989: 97–8). Yet, it may be argued, the effected partition of Newton’s law does not consist of independently acceptable generalizations. In order, however, to judge this, we first need to have a clear understanding of what it is for a law to be acceptable independently of others. Friedman’s criterion – that a law L_1 is acceptable independently of law L_2 if there are sufficient grounds for accepting L_1 that are not also sufficient grounds for accepting L_2 – admits of two readings, one historical and another conceptual.

On a historical reading, whether a generalization is acceptable independently of another will depend on the *order* in which the

scientific community placed the two generalizations in the body *K* of accepted regularities. If, for instance, Newton's law of gravity was, as a matter of historical fact, accepted *after* (and because of) at least one of the three foregoing generalizations, then it would *not* count as independently acceptable. Hence it would not count as atomic. So it would not explain, by unification, these three regularities as well as Kepler's laws, Galileo's laws, and so on. But then the question is this: why should the order in which a law was accepted have any bearing on its explanatory power? I see no motivation for this. The explanatory power of a law seems to be an objective matter; hence it ought to be independent of the order in which this law came to be accepted.

The other option is to go for a conceptual reading of *independently acceptable*: the acceptance of a generalization is *conceptually* independent of the generalizations that form its partition. If, however, we endorsed this reading, it would turn out that a mere conjunction can be atomic and hence explanatory of its conjuncts. For instance, the acceptance of the conjunction of Boyle's law with the law of adiabatic change could be conceptually independent of the two individual conjuncts, in that one could (conceptually) accept first the conjunction, and then each conjunct. This would render the conjunction atomic, and hence explanatory of Boyle's law. So there is a tension in Friedman's account. Given that any lawlike statement can be partitioned into a set of other statements, whether this lawlike statement is explanatory will depend on whether it is acceptable independently of the members of the partition. If we go for a historical understanding of the notion of *independently acceptable regularities*, then some genuinely explanatory laws might *not* count as explanatory. But if we go for a logical understanding of *independently acceptable regularities*, then some unwanted conjunctions may be atomic, and hence genuinely explanatory.

There is another retort that the advocates of "atomicity" might favour. They might insist that not all syntactic partitions of a lawlike statement will undermine its atomicity, since not all syntactic partitions will correspond to *natural kind* predicates. So the thought might be that whereas *F* and *G* in "All *F*s are *G*s" are *natural kind* predicates, the predicates, *F* & *H* and *F* & *not-H*, which can be used to form the logically equivalent partition {All (*F*s

& *Hs*) are *Gs*; All (*Fs* & *not-Hs*) are *Gs*}, are not necessarily *natural kind* predicates. This issue is discussed in some detail by Salmon (1989: 96–8). Whichever way one looks at it, it shows some important weakness of Friedman’s approach. In order to be viable, this approach requires a theory of what predicates pick out natural kinds. This cannot be a purely syntactic matter. One standard thought has been that the predicates that pick out natural kinds are the predicates that are constituents of genuine lawlike statements. But on Friedman’s approach it seems that this thought would lead to circularity. In order to say what statements are genuinely atomic, and hence what statements express explanatory laws, we first need to show what syntactically possible partitions are *not* acceptable. If we do that by means of a theory of what predicates pick out natural kinds, then we cannot, on pain of circularity, say that those predicates pick out natural kinds that are constituents of statements that express explanatory laws. This last objection, however, may not be as fatal as it first sounds. The genuine link that there is between delineating what laws of nature are and what kinds are *natural* has led many philosophers to think that the two issues can only be sorted out together. The concept of a law of nature and the concept of a natural-kind predicate form a family: one cannot be delineated without the other.

I think the basic flaw in Friedman’s approach is the following. He defines unification in a syntactic fashion. In this sense, he’s very close to the original Hempelian attempt to characterize *explanation* in a syntactic manner. As we saw in section 8.2, Hempel ran into the problem of how to distinguish between genuine laws and merely accidentally true generalizations. Purely syntactic considerations could not underwrite this distinction. Friedman attempted to solve this problem by appealing to unification. But the old problem reappears in a new guise. Now it is the problem of how to distinguish between *good* unifiers (such as Newton’s laws) and *bad* unifiers (such as mere conjunctions). Yet a purely syntactic characterization is doomed to fail, no less than it failed as a solution to Hempel’s original problem.

Having said all this, there may be a way to defend the spirit of Friedman’s model of unification. But the model should be non-syntactic. I want to motivate my suggestion by noting that the basic idea behind Friedman’s programme is right: explanation *does*

amount to a reduction of the number of independently acceptable laws. The crux of this idea is that all apparently independently acceptable regularities are consequences of more basic laws. But if syntactic criteria such as *atomicity* are bound to fail, what could take their place? My suggestion is that a law should count as unifying, and hence explanatory, if it has *excess content* over the regularities it unifies, where this excess content should be understood in terms of *hitherto unknown regularities*. That is, a lawlike statement is genuinely unifying of a number of regularities if, apart from entailing them, it also entails *novel regularities*. Such novel, that is, hitherto unforeseen, regularities cannot possibly be acceptable independently of the unifying law that predicted them. For their presence in nature is suggested by the unifying law. It is the higher-level law that predicts them that makes them, as it were, available and acceptable. This criterion is clearly not syntactic, since what regularities should count as novel cannot be settled by any syntactic criteria. Besides, it is easy to see that, on the presently suggested account, all trivial conjunctions of existing laws will fail to be explanatory, since they necessarily fail to predict novel regularities. If this suggestion is on the right track, then it transpires that unification and the ability to predict novel regularities are the two sides of the same coin. They constitute what is sometimes called explanatory power: novel prediction and genuine explanation are brought together under the same banner.

10.3 Unified explanatory store

The failures of Friedman's approach to unification have led Kitcher (1981) to advance an alternative view, which changes substantially the characterization of unification. In his earlier work, Kitcher was not so much concerned with the nature of laws of nature. He was mostly concerned with the explication of the concept of unification. The key idea, however, is a species of the empiricist project to tie the concept of explanation to the concept of the deductive systematization. He calls us to envisage a set K of statements accepted by the scientific community. K is consistent and deductively closed. An "explanatory store $E(K)$ " over K is "the best systematisation of K " (1981: 337). The best systematization, however, is not what Friedman took it to be. It is not couched in terms of the minimal set

of lawlike statements that need to be assumed in order for the rest of the statements in K to follow from them. For Kitcher, the best systematization is still couched in terms of the derivation of statements of K that best unifies K , but the unification of K is not taken to be a function of the size (cardinality) of its set of axioms. Rather, Kitcher takes unification to be a function of the number of explanatory patterns, or schemata, that are necessary to account for the statements of K . The smaller this number is, the more unified is $E(K)$. Given a small number of explanatory patterns, it may turn out that the number of facts that need to be accepted as brute in the derivations of statements of K might be small too. So it may be that Kitcher's unification entails (the thrust of) Friedman's unification. But it is important to stress that what bears the burden of unification for Kitcher is the explanatory pattern (schema). As he put it: "Science advances our understanding of nature by showing us how to derive descriptions of many phenomena, using the same pattern of derivation again and again, and in demonstrating this, it teaches us how to reduce the number of types of fact that we accept as ultimate" (1989: 432).

10.3.1 Explanatory schemata

Before we analyse further Kitcher's central idea, we need to understand his notion of an explanatory schema (or pattern). To fix our ideas, let us use an example (cf. Kitcher 1989: 445–7). Take one of the fundamental issues in the post-Daltonian chemistry; namely, the explanation of the fact that the compounds of X and Y always contains X and Y in the weight ratio $m:n$. Kitcher suggests that Dalton's approach can be seen as involving the following explanatory schema.

1. The compound Z between X and Y has an atomic formula of the form: $X_p Y_q$.
2. The atomic weight of X is x and the atomic weight of Y is y .
3. The weight ratio of X to Y is $px:qy$ ($= m:n$).

This schema can be repeatedly (and successfully) applied to many cases of compounds. Take Z to be water. Then (1) $X = \text{H}$ (hydrogen) and $Y = \text{O}$ (oxygen) and Z is H_2O_1 . Then (2) $x = 1$ and $y = 16$.

Then (3) $(2 \times 1):(1 \times 16) = 2:16 = 1:8 (= m:n)$. The structure of this explanatory schema (general argument-pattern) is an ordered triple: <schematic argument, filling instructions, classification>.

- The *schematic argument* is (1) to (3) above. It is schematic because it consists of schematic sentences. These are sentences in which some non-logical expressions occurring in them (e.g. names of chemical elements) are replaced by dummy letters (e.g. Z, X, Y), which can take several values.
- The *filling instructions* are directions for replacing the dummy letters of the schematic sentences with their appropriate values. In the example at hand, the dummy letters X and Y should be replaced by names of elements (e.g. hydrogen and oxygen), the dummy letters p and q should take natural numbers as values, and the dummy letters x, y should take real numbers as values.
- The *classification* is a set of statements that describe the inferential structure of the schema. In the case at hand, the classification dictates that (1) and (2) are the premises of the argument while (3) is the conclusion.

Explanatory schemata are the vehicles of explanation. The explanatory store $E(K)$ is “a reserve of explanatory arguments” (Kitcher 1981: 332), whose repeated applications to many phenomena brings order – and hence unifies – K . Kitcher (1981: 333) sums up his position thus: “a theory unifies our beliefs when it provides one (more generally, a few) pattern(s) of argument which can be used in the derivation of a large number of sentences we accept”. Among the many possible systematizations of a body K of beliefs, the one that should be accepted as the *best*, that is, the most unified one, is the one that rests on the fewest possible argument-patterns whose repeated application is enough to derive the greatest number of sentences of K . So Kitcher’s approach is a variant of the *best system* approach. In fact, it explains what it is for a systematization to be *best*. The unifying quality of the system is to be a function of the patterns of derivation in the system: the fewer the number of patterns and the larger the number of sentences of K that are the conclusions of instances of these patterns, the more unified the system is.

A central thought in Kitcher’s account is that explanations are arguments, and in particular *deductive* arguments. The best

systematization is still a deductive systematization, even if what effects the systematization is the number of deductive patterns that are admissible, and not the number of axioms of the *best system*. In this sense, Kitcher's approach is a descendant of Hempel's DN model. It shares some of its most important features and consequences. The relation of explanatory dependence is a relation between sentences and it should be such that it instantiates a deductively valid argument with (a description of) the *explanandum* as its conclusion. As Kitcher (1989: 431) put it: "the systematisation approach retains the Hempelian idea that to explain a phenomenon is to produce an argument whose conclusion describes the phenomenon . . .". Yet we need to be careful here. Kitcher's account, as it now stands, does *not* demand that the premises of explanatory arguments be laws of nature. It does not even demand that they be universally quantified statements. They may be, and yet they may not. So, as it stands, Kitcher's account need not be a way to explicate what the laws of nature are. Nor does it demand that all explanation be *nomological*.

However, it seems that statements that express genuine laws of nature are uniquely apt to do the job that Kitcher demands of explanation. By being genuinely lawlike, these statements can underwrite the power that some schemata have to be repeatedly employed in explanations of singular events. Take the case, discussed also by Hempel, of trying to explain why John Jones is bald. As we have already seen in section 8.2, Hempel rightly thought it inadmissible to explain this fact by constructing a DN argument whose premises are the following: "John Jones is a member of the Greenbury School Board for 1964" and "All members of the Greenbury School Board for 1964 are bald". His reason, as you may recall, was that the statement "All members of the Greenbury School Board for 1964 are bald" did not express a genuine law. Kitcher agrees with Hempel that this explanation is inadmissible: it rests on an accidentally true generalization. But how is he to draw the distinction between laws and accidents within his own account? He says that an argument-pattern that aims to explain why certain individuals are bald by employing the sentence "All members of the Greenbury School Board for 1964 are bald" is not "generally applicable" (Kitcher 1981: 341). On the contrary, an argument-pattern that would aim to explain why certain individuals are bald by reference to some principles of physiology would be generally applicable.

What, however, underwrites the difference in the applicability of argument-patterns such as the above is that the former rests on an accidental generalization while the latter rests on genuine laws. It's not just that "All members of the Greenbury School Board for 1964 are bald" has a finite number of instances – a fact that would impair its applicability. Kepler's first law has only a finite number of instances, and yet we think that its presence in an argument-pattern would not impair its applicability. So Kitcher needs to tie the explanatory applicability of an argument-pattern to the presence of genuine lawlike statements in it. He is certainly willing to exclude accidental generalizations from featuring in explanatory arguments (1981: 341), but he then needs to explain what distinguishes between laws and accidents. In a later piece, Kitcher takes very seriously the thought that the statements that express laws "are the universal premises that occur in explanatory derivations" (1989: 447). Still, it is not clear what distinguishes between laws and accidents in his approach. He does draw a distinction between "mini-laws" and "maxi-laws", where the former can be seen as low-level generalizations (e.g. that sodium and chlorine combine in a one-one ratio), while the latter can be seen as high-level generalizations (e.g. Mendel's laws in genetics). These are different in their scope: "mini-laws" are less universally applicable than "maxi-laws". Both are laws, however, since they occur in patterns of derivation of (descriptions of) the phenomena to be explained. We are not told, though, what exactly it is that makes them laws, as opposed to accidents.

10.3.2 Explanatory asymmetries

Admittedly, Kitcher's approach fares much better in its attempt to explain the *de facto* asymmetric character of good explanations. The reader might recall from section 8.5 that one of the problems that plagued Hempel's approach was brought to light by Bromberger's counter-example of the height of the flagpole and the length of the shadow. The moral of this counter-example was that the asymmetry between an explanation of the length of the shadow in terms of the height of the pole and the explanation of the height of the pole by the length of the shadow cannot be adequately captured by Hempel's DN model. How does Kitcher make sure that the intuitive asymmetry in the order of explanation is restored?

Working within the framework of argument-patterns, Kitcher must show that an argument-pattern that explains the height of the pole in terms of the length of the shadow is inadequate. And that's what he does. He takes the class of what he calls "origin and development derivations" (1981: 341) and argues that they specify an argument-pattern that is to be preferred over an alternative "shadow-pattern". According to the "origin and development pattern", we explain the dimensions of a thing by reference to the conditions under which it was designed.³ So we explain the height of the flag-pole by reference to the circumstances that led to the formation of the pole. This argument-pattern is generally applicable to, and variably instantiated in, many cases. In particular, it can be instantiated in cases where an object casts no shadow. The "shadow-pattern", however, which explains the dimensions of objects with reference to the lengths of the shadows they cast, is not similarly generalizable and variably applicable. It cannot, for instance, be applied to cases of objects that cast no shadow. So Kitcher thinks that the "origin-and-development" argument-pattern is more unifying of our body of knowledge K . The explanatory store $E(K)$, which includes this pattern, is simpler than the one that includes the "shadow-pattern". After all, if $E(K)$ were to include the "shadow-pattern", it would also have to include the "origin and development" one, which would be applicable to cases where objects cast no shadow.

A natural objection at this point would be that the "shadow pattern" can be easily modified in order to be more generally applicable. Since objects have a *disposition* to cast a shadow under certain circumstances of illumination, the "shadow pattern" could become a "dispositional shadow pattern", according to which we explain the height of an object not by reference to the length of its actual shadow but with reference to the length of the shadow it *would* cast under certain circumstances. Considering this objection, Kitcher (1989: 486) rightly says that the "dispositional shadow pattern" is still less satisfactory than the "origin and development pattern", for it's more complicated. If we followed the "dispositional shadow pattern", then, when, for instance, it comes to transparent objects, we would have to credit them with one more disposition, namely, the disposition to be coloured, which alongside the disposition to cast a shadow, would explain their height in terms of the length of *the shadow they would cast if they were not transparent*.

Still, there seems to be some room for dissatisfaction with Kitcher's account of the asymmetries in explanation. One could think that the very fact that the unifying argument-patterns are those that preserve the intuitive asymmetry in the order of explanation points to a deeper characteristic they have; namely, that they capture facts about the *causal order* of the world.⁴

Why would this thought be unappealing to Kitcher? Being within the Humean camp, Kitcher wants his theory of explanation to capture the empiricist idea that the causal order of the world does not ontologically precede the explanatory order, as the latter is described in the most unified theoretical system of the world. He (1989: 149) views his own account as falling within the Hempelian legacy, which takes causal notions to be "understood either in terms of the concept of explanation or in terms of concepts that are themselves sufficient for analysing explanation". Hence he takes as his task to offer a theory of explanation that makes clear that "one event is causally dependent on another just in case there is an explanation of the former that includes a description of the latter" (1989: 420). He is then pressed to explain some salient features of explanation, in particular the fact that it can display asymmetries between the *explanans* and the *explanandum*, in non-causal terms. The order of dependence between explanation and causation will be the central issue to be discussed in the final chapter of this book.

But before we turn our attention to it, the reader might wonder whether Kitcher's account of unification in terms of argument-patterns is satisfactory. The notion of an argument-pattern is clear enough and does seem to capture some sense in which a system is unified. But when argument-patterns are applied to several cases, things seem to be more complicated than Kitcher thinks. Take one of his own examples: Newton's second law of motion. Once we are clear on the notion of *force*, Newton's law $F = ma$ can be seen as specifying a Kitcher-like argument-pattern. The whole problem, however, is that none of the elements of the triple that specify an argument-pattern – namely, schematic argument, filling instructions, classification – can capture the all-important concept of a force-function. Each specific application of Newton's law requires, as Cartwright has repeatedly stressed, the prior specification of a suitable force-function. So, when we deal with a pendulum, we need to introduce a different force-function (e.g. $F = -Kx$) than

when we are faced with a planet revolving around the sun. It's not part of the schematic argument what force-functions are applicable. Nor can this be added to the filling instructions, simply because the force-functions may be too diverse, or hitherto unspecified.

There is clearly something to the idea that, given a repertoire of force-functions, Newton's second law can be schematized *à la* Kitcher. But part of explaining a singular event is surely to figure out *what* force-function applies to this particular case. Besides, even when we have chosen the relevant force-function, we need to introduce further assumptions, related to the specific domain of application, which will typically rest on idealizations. All these cannot be part of the explanatory pattern. What really seems to matter in most (if not all) cases is that the phenomena to be explained are traced back to some kind of basic law, such as $F = ma$. It's not so much that we can repeatedly apply a certain argument-pattern to derive more specific cases. Instead, more typically, we show how specific cases can be reduced to being instances of some basic principles. That these basic principles will be applicable to many phenomena follows from their universal character. But it seems irrelevant whether or not the repertoire of the arguments from which (descriptions of) several phenomena derive is small or large. Unification consists in minimizing the number of *types* of general principles, which are enough to account for the phenomena. Admittedly, this view is closer to Friedman's than to Kitcher's. But so be it.

The metaphysics of

11 explanation

It's about time to tackle an important issue that lies behind the debate around the nature of explanation. It is, I think, the basic *metaphysical* issue: what comes first, explanation or causation? Kitcher's and Salmon's general views will be the focal point of this chapter.

11.1 "Bottom-up" vs "top-down"

Salmon has made a distinction between three broad approaches to the nature of explanation. He has called them the "epistemic conception", the "modal conception" and the "ontic conception".

The *epistemic conception* is the Hempelian conception. It makes the concept of explanation broadly epistemic, since it takes explanation to be, ultimately, nomic expectability. The *modal conception* differs from the epistemic mostly in its account of necessity. The *explanandum* is said to follow necessarily from the *explanans*, in the sense that it was *not* possible for it not to occur, given the relevant laws. Von Wright (1971: 13) put the point thus: "What makes a deductive-nomological explanation 'explain' is, one might say, that it tells us why E *had* to be (occur), why E was *necessary* once the basis [body of explanatory facts] is there and the laws are accepted."¹

The *ontic conception*, the one advocated by Salmon (1984), takes explanation to be intimately linked to *causation*. As he (1984: 19) explains: "To give scientific explanations is to show how events . . . fit into the causal structure of the world." Salmon takes the

world to have an already built-in causal structure. Explanation is then seen as the process in virtue of which the *explananda* are placed in their right position within this causal structure.

This view might sound acceptable to all sides of the debate. But what matters here is the order of dependence. Where the epistemic conception has aimed to capture relations of causal dependence by means of relations of explanatory dependence, the ontic conception does the opposite. This difference is highlighted if we consider the issue of laws of nature. Salmon is, to a certain extent, a Humean. He (1984: 121) thinks that the laws of nature are, ultimately, regularities. Yet he also thinks that not all regularities have explanatory power. This view might be seen to coincide with the distinction between laws and accidents, which was discussed in Chapter 5. But in fact it cuts deeper. Salmon looks for a distinction between causal laws and non-causal laws. The law of ideal gases, for instance, is “a lawful regularity”. Yet, Salmon stresses, it is *not* a causal law, since it does not display an underlying causal mechanism by virtue of which pressure is causally connected to the other macroscopic parameters of gases (temperature and volume). The laws of the molecular-kinetic theory of gases, on the other hand, are causal laws: for they provide causal mechanisms. In fact, Salmon argues, it is these laws that genuinely explain the law of ideal gases. Hence, whatever explanatory import the latter law might have is parasitic on being itself explained by the causal laws of the molecular-kinetic theory of gases. The task then is to explain what makes a law causal. We have already dealt with Salmon’s theory of causal processes in Chapter 4. So it is enough to remind the reader that, for Salmon, causal regularities are those regularities that are underpinned by causal processes or mechanisms and are involved in causal interactions. The issue that concerns us here is the broad metaphysical picture that Salmon paints. According to Salmon (1989: 128) “explanatory knowledge is knowledge of the causal mechanisms, and mechanisms of other types perhaps, that produce the phenomena with which we are concerned”.

Kitcher (1985: 638) has rightly called Salmon’s approach “bottom-up”. This approach takes causal relations to be *prior* to relations of explanatory dependence. What explains what is parasitic on (or determined by) what causes what. So we should first discern causal relations among particular events, and then conceive of the

task of explanation as identifying the causal mechanisms that produce the events for which we seek an explanation. To this approach, Kitcher contrasts a “top-down” one. We begin with a “unified deductive systematisation of our beliefs”. Then we proceed to make ascriptions of causal dependencies (i.e. of relations of cause and effect), which are parasitic on (or determined by) the relations of explanatory dependence that emerge within the best unified system. “On this approach”, Kitcher (1985: 639) notes, “theoretical explanation is primary. Causal concepts are derivative from explanatory concepts”.

We should try to be clear on what really is at issue between these two approaches. Unification is *not* at issue here. Salmon is as willing as Kitcher is to adopt unification as the goal of scientific explanation. He (Salmon 1985: 651) takes it that an advocate of a “mechanistic” view of explanation (namely, of the view that explanation amounts to the identification of causal mechanisms) is perfectly happy with the idea that there is a small repertoire of causal mechanisms that work in widely different circumstances. He is also perfectly happy with the view that “the basic mechanisms conform to general laws” (*ibid.*). Unification, Salmon stresses, promotes our understanding of the phenomena, irrespective of whether one takes the “bottom-up” or the “top-down” line. So what is at stake in this debate? There is a narrow and a broad issue at stake. The *narrow* issue concerns the commitment of the “top-down” approach to explanation as a species of *deductive derivation*. The *broad* issue, which is far more important from a metaphysical point of view, concerns the role of causation in explanation. In particular, it concerns the fundamental metaphysical question: what comes first, explanation or causation? I’ll leave the broad issue for the final section 11.3. So let’s try to cast some light on the narrow issue.

11.2 Deductive chauvinism

The commitment to explanation as a species of deductive derivation has been so pervasive that it can hardly be exaggerated. It featured prominently in Hempel’s deductive–nomological model of explanation (see Chapter 8) and it reappeared under a new guise in both Friedman’s and Kitcher’s accounts of explanation as unification (Chapter 10). In a different form, it also appeared in the MRL view of laws (see Chapter 5). The charge that Salmon (1985:

652) has levelled against this *derivationist* approach to explanation is that it leaves unaccounted for the role of statistical regularities in the explanation of singular events. This might sound odd, given the fact that, as we already saw in Chapter 9, Humeans have tried hard to provide a model of statistical explanation. But Salmon's point is that a derivationist approach to explanation would necessarily leave out some important *causal* facts, which are not expressible within the best unified *deductive* system of the world.

As a relevant example, let us think of the case that Salmon presents. Suppose that a Geiger counter registers 99 per cent of the impinging photons. Suppose also that if it registers something – if, say, it makes a click – then there is always a photon present. In other words, suppose that although it might fail to register 1 in 100 photons, it is only photons that it registers. So the presence of a photon is a necessary but not sufficient condition for a click of the counter. There is a strong intuitive sense in which we can say that a certain click of the counter was *caused* by the impingement of a photon, and hence that it was the impingement of the photon that explains the click. Yet we cannot devise a deductively valid argument to explain this event. (A description of) the *explanandum* (i.e. the click) cannot be the conclusion of a deductively valid argument, since the *explanans* do not contain a universal lawlike statement. Even if it is a law that 99 per cent of photons are detected by the Geiger counter, we cannot deduce from it (together with other antecedent conditions) that a click *will* occur. The moral we are invited to draw from this example is that some genuine causal facts (e.g. that the photon *caused* the click of the counter) are unwarrantedly left out of the picture, if we stick to a derivationist account of explanation.

How do derivationists react to these counter-examples? Kitcher is ready to accept the charge of “deductive chauvinism”. (Although the charge comes from Salmon, the expression comes from Coffa.) As Kitcher says: “*In a certain sense*, all explanation is deductive” (1989: 448). That *some* statistical regularities can be explained by means of suitable deductive arguments is no news, of course. As we have seen in section 9.1, it was to this purpose that Hempel introduced the DS model of explanation. An advocate of the “top-down” approach to explanation, such as Kitcher, can accept that the best unified system of the world can have argument-patterns

that involve statistical regularities. If all statistical explanation were of the DS variety, then Kitcher's thesis that all explanation is deductive would be home free.

Unfortunately for the derivationalist there are legitimate forms of statistical explanations that cannot take the guise of a statistical deduction. Such explanations are very prominent in the so-called special sciences (economics, psychology, sociology, etc.), but they also seem essential in some part of physics – especially in quantum mechanics. In section 9.2, we saw Hempel's attempt to offer an IS model for the explanation of singular events that happen with probability less than one. But Kitcher has a hard-line reaction to the problem of singular statistical explanation. In effect, he *denies* that there are irreducibly singular statistical explanations. If all explanation is to be deductive, then all apparently irreducible statistical explanations must be seen as “placeholders” for “underlying, unknown, deductive” explanations (1989: 499). The idea then is that whenever we are confronted with a statistical explanation of a singular event, there are further facts such that, were they known, they could be used to afford a fully deductive explanation of the occurrence of the singular event in question. So, Kitcher says, singular statistical explanations express our *ignorance* of the further factors that fully determine the occurrence of the *explanandum*. He adds that when the probability that the *explanandum* will occur is high, we have *evidence* that there are such further factors to be discovered.

There is no doubt that Kitcher's view is consistent. Still, don't we feel that there is space for genuine statistical explanation of singular events, especially if indeterminism is true? Salmon, for instance, argues that (a) we *should* raise the question of why an individual event, which is governed by a statistical law, did happen; and (b) we should explain not only the events that have high probability of happening but also those with a low probability. This is made possible in Salmon's theory of explanation. Whether determinism or indeterminism is true, we can explain an individual event by fitting it into a pattern, “one which is constituted by universal or statistical regularities in the world” (1984: 119). If determinism is true, then ultimately there are no statistical or stochastic regularities. But if indeterminism is true, then Salmon's thought becomes really interesting. To fit an event-type *E* into a statistical (stochastic) pattern is

to say that there are circumstances C such that they yield the event-type E in a certain percentage r of cases; and the very same circumstances C are such that they do not yield the event-type E in a percentage $1-r$ of cases. The very same circumstances C are responsible for the occurrence of E in some cases and its failure to occur in some others. If indeterminism is true, there are no further facts that, were they taken into account, would further explain (determine) whether E or *not- E* will occur. But Salmon thinks this is not a defect. It is simply a reflection of the fact that indeterminism might be true. His point is that there can be genuine explanation, even if indeterminism is true (cf. 1984: 120).

Salmon warns us *against* a condition on explanation, which has been put forward by advocates of determinism. Salmon (1984: 113) puts this condition thus: "If a given set of facts provides an adequate explanation of some event E , then those same facts cannot suffice to explain the nonoccurrence of E ." It is precisely this adequacy condition that Salmon doubts. His point is that the very same set of facts *can* be employed to explain why E occurred on some occasion, but didn't occur on some other. This is done, he argues, by laying bare the "physical mechanisms" that produce occasionally E and occasionally *not- E* . These mechanisms will be stochastic, since they produce an effect on some occasions and another effect on some others.

I am not sure, however, that there is genuine disagreement here between Kitcher and Salmon. For although it is true that the occurrence of the individual event can be placed in a stochastic pattern, we still seem to *lack* an *explanation* of why this individual event occurred. Kitcher would be happy with this lack, since he thinks that what we seem to lack, we shouldn't seek anyway. But has Salmon offered an explanation, or has he simply also stated that we shouldn't seek for a further explanation of the individual occurrence? It's useful here to recall Railton's reaction to this problem – see section 9.4. He just added to a DNP argument the *addendum*: *and the unlikely event did happen*. But in what sense is this an explanation of its occurrence? To put the point differently, it seems perfectly possible for a derivationist to say the following: we can certainly deduce the lawful probability that a certain chance event happens, for example, the decay of a radioactive atom (a thing that would amount to a fully deterministic explanation of a probability

state). But we don't thereby explain why this particular event happened on a particular occasion – we just admit that it did. In fact, we can say no more. We can accept with “natural piety” that events that only have some probability of happening (i.e. chance events) *do* happen. We just have to wait to see how things turn out.

Returning to this issue in a later piece, Salmon invited us to perform a “gestalt switch” and to abandon our “widely held intuitions” that we should seek a further story as to why this rather than that stochastic event happened (cf. 1997a: 328–9). True, our deterministic intuitions might be in need of reform. But it may be that our concept of explanation is such that it cannot really cover genuine stochastic singular events. Shall we then reform our intuitions, or shall we reform our concept of explanation? I don't yet know the answer to this question, if there is any.

11.3 What comes first, explanation or causation?

Let's now return to the broad issue that separates the “top-down” and the “bottom-up” approach to explanation. As noted above, this issue concerns the fundamental metaphysical question: what is primary, explanation or causation?

Recall Kitcher's view of explanation, which was discussed in section 10.3. As he admits, he has only offered an account of “acceptable explanations”, that is, an account of “what conditions must be met for a derivation to be an acceptable explanation of its conclusion relative to a belief corpus *K*” (1989: 494). But acceptable explanations may fail to be *correct* explanations. This failure might be due to the fact that explanations might not capture the causal structure of the world. One way to cash in the idea of the causal structure of the world is to say that there is an objective fact of the matter as to the causal connections that hold between event-types in the world. If this is so, then an appeal to the causal structure of the world could be the arbiter of claims about the correctness of explanations.

Appealing though this claim may be, it is not available to Humeans who, like Kitcher, endorse the view that the causal order of the world emerges out of relations of explanatory dependence in the unified deductive systematization of our beliefs. So how can Humeans proceed in order to ground the thought that there is still

space to distinguish between correct and incorrect explanations?

One first step is to deploy a device akin to the main idea behind the MRL view of laws of nature (see section 5.6). We don't have to tie the notion of explanation to what we *currently* accept as true; nor to our *present* beliefs. We can envisage what Kitcher calls an "ideal" explanatory store. This is "the complete set of truths about nature" (1989: 495). Kitcher calls this store "an ideal Hume corpus". It is a *Hume* corpus because, in the Humean-empiricist spirit, the truths expressed in this corpus do *not* involve "causal, explanatory or counterfactuals concepts" (1989: 495). So an ideal Hume corpus is best suited to do two jobs: first, to deflect the charge that what we now believe is too parochial to capture correct explanation; and secondly, to guarantee that no irreducible causal claims are sneaked into the characterization of explanation. Once this corpus is envisaged to be fixed, then all relations of explanatory dependence – and hence all explanations – are fixed as well: they consist in those relations of deductive entailment among statements of the ideal Hume corpus that best unify the ideal Hume corpus. So, says Kitcher (1989: 495), to say that *F* is *explanatorily relevant* to *P* is to say: "there is a derivation of *P* that belongs to the best unifying systematisation of an ideal Hume corpus such that there is a premise of the derivation in which reference to *F* is made". Add to this the basic empiricist thesis: "If *F* is causally relevant to *P* then *F* is explanatorily relevant to *P*" (*ibid.*). What we then get is the view that all relations of causal dependence are fixed (captured) by the relations of explanatory dependence that appear in the ideal Hume corpus. So saying that *C* causes *E*, or that *C* causally explains *E*, becomes a shorthand for a more complex construction: one that consists in a derivation, in the ideal Hume corpus, of (a description of) *E* from an argument, one of its premises being (a description of) *C*. The modern Humean project seems to be completed: causation mirrors explanation in an ideal Hume corpus, where the latter is understood in non-causal terms. What is gained by this move? If Kitcher is right in his claim that "the 'because' of causation is always derivative from the 'because' of explanation" (1989: 477), then Humeans can, after all, succeed in showing (a) that ascriptions of causal relations among events can be grounded in non-causal facts, and (b) that causal (that is, explanatory) knowledge of the world is possible.

The problem, however, is that Kitcher's approach clashes with an important intuition we have. Kitcher expresses it thus:

It is possible (and may, for all we know, be true) that there is a factor *F* that is causally relevant to some phenomenon *P* such that derivation of a description of *P* belonging to the best unifying systematisation of an ideal Hume corpus would not contain any premise making reference to *F*. (1989: 595)

Let's call this the *possibility of divergence*.

The possibility of divergence: There may be causal facts (or relations of causal dependence) that are not captured by some explanatory derivation in the ideal Hume corpus.

If Kitcher is to be consistent, then he must *block* this possibility.

11.3.1 A Socratic dilemma

One important consequence of the issue we are currently discussing is this. As we saw towards the end of section 11.1, Salmon is perfectly happy with the idea that explanation amounts to unification in the sense that there are only a few underlying causal mechanisms upon which we depend for explanation. But Salmon is also perfectly happy with the *possibility of divergence*. In fact, he must leave this possibility open in order to claim that the causal order of the world is (metaphysically) prior to the explanatory order. So, for Salmon, it is at most a *contingent* truth that the unified explanatory order and the causal order coincide. That is to say, it may well be as a matter of fact true that the ideal unified deductive-explanatory system of the world reflects the pre-existing causal order of the world. But the possibility of divergence is open in the sense that it is *not* a necessary truth that the ideal explanatory order and the causal order have to coincide. Kitcher, on the other hand, must make this coincidence necessarily true. By denying the *possibility of divergence*, he must accept that there cannot be causal facts that are not captured in the ideal Hume corpus. In other words, Kitcher must make it *constitutive* of the concept of explanation he advances that it captures (or exhausts) the causal order of the world.

The issue between Salmon and Kitcher then can be seen as an instance of the Socratic *Euthyphro contrast*. Socrates asked Euthyphro to take sides on the following dilemma: “whether the pious or holy is beloved by the gods because it is holy, or holy because it is beloved of the gods”. The contrast is not particularly theological; it is very general. It invites us to see the order of dependence among some things. Paraphrasing Socrates, one could ask: are some events (or facts, if you will) explanatorily related because they are causally related, or are some events (or facts) causally related because they are explanatorily related? The issue here is not extensional. Pious acts might well be just those acts that are beloved by the gods. Similarly, the causal relations in the world might be just those relations that are captured by the explanatory dependencies issued by the ideal Hume corpus.² The issue, then, concerns the *order of dependence*.

In the original Euthyphro contrast, Socrates invited Euthyphro to see that it made a difference which side of the contrast one took as primary. In particular, Socrates suggested that it is an independent fact that some acts are pious, which might be *evidenced* by the fact that gods love them. It is not, that is, the fact that gods love some acts that makes these acts pious. Piety is not constituted by the fact that gods love some actions and not others. Rather, gods love some actions *because* they are pious. Similarly, in the paraphrased Euthyphro contrast, Salmon suggests that it is an independent fact that some events are causally related, which is *evidenced* by the fact that they are also related by some relation of explanatory dependence in the ideal Hume corpus. Kitcher, on the other hand, takes Euthyphro’s side: what constitutes the relation of causal dependence is the fact that the events that are said to be causally related are already explanatorily related in the ideal Hume corpus. The Euthyphro contrast brings to focus the core metaphysical issue that separates the two competing approaches to explanation.

So how does Kitcher justify his preferred side on the Euthyphro contrast? In particular, how does he block the *possibility of divergence* between the causal order and the explanatory order, which *should* be blocked if his preferred side on the Euthyphro contrast is to be defended? What Kitcher says is interesting but, ultimately, problematic. He rejects the *possibility of divergence* by introducing a notion of the “limit of the rational development of scientific

practice" (1989: 498), and by arguing that if we were to reach this limit, there would be "no sense to the notion of causal relevance independent of that of explanatory relevance" and "no sense to the notion of explanatory relevance except that of figuring in the systematisation of belief in the limit of scientific inquiry" (1989: 499). Yet it's not clear that Kitcher has done anything other than to just dig in his heels by *denying* the *possibility of divergence*.

Even if we envisage the limit of inquiry, there might not be just one way to best unify the ideal Hume corpus. The reader will recall that a similar problem cropped up in our discussion of the MRL approach to laws (see section 5.6). The prime concern, there, was to distinguish between genuine laws and accidentally true generalizations. Genuine laws, we saw, are those whose descriptions belong to the best deductive systematization of our knowledge of the world. But, as it was pointed out there, there is an important objection to this view: what statements express genuine laws will depend on the *way* our knowledge of the world is organized into a deductive system. Hence, what the laws of nature are is not a fully objective matter. The same objection may be raised against Kitcher's own subsumption of causal order under explanatory order. Intuitively at least, causal order is objective: what causes what is an objective feature of the world. But if causal dependencies are to mirror explanatory dependencies, and if the latter are a matter of organizing our knowledge of the world into a best unified deductive system, then causal relations will not be objective. Why? Because the relations of explanatory dependence will not be objective: they will depend on the way the ideal Hume corpus is organized. It is indeed possible that there may be different, even incompatible, ways to organize this corpus. If we were to assume that there is an objective causal structure of the world, and that relations of explanatory dependence should mirror this objective causal structure, then, at least, we would have an *external standard* to judge the correctness of the ensued explanatory order: the best unified deductive systematization of the ideal Hume corpus is the one that characterizes correctly the objective causal structure of the world. Yet, if we follow Kitcher, there cannot be such an external standard: the correct causal order just *is* the ideal explanatory order. What, however, determines the correctness of this explanatory order? Briefly put, the objection to Kitcher's move is that it makes causation mind-dependent.

11.3.2 A Ramseyan finale

Humeans have two options. The first is to accept the charge of mind-dependence. This route is followed by Kitcher (1989), when he chooses to dismiss the possibility of a divergence between the causal order of the world and the explanatory order imposed by the best unified ideal Hume corpus. He simply takes the neo-Kantian line that the causal order is imposed on the world by our best unified ideal Hume corpus. He says: "I recommend rejecting the idea that there are causal truths that are independent of our search for order in the phenomena. Taking a clue from Kant and Peirce, we adopt a different view of truth and correctness, and so solve the problem with which we begun" (1989: 497). But the price we are invited to pay, if we follow this option, is very high. The problem is not just that Kitcher's preferred answer slides towards (transcendental) idealism.³ This is already too much, but the thought may be that since Humeans have to engage in metaphysics after all, they might opt for idealist metaphysics. The real problem lies elsewhere. By blocking a priori the *possibility of a divergence* between the causal order of the world and the explanatory order imposed by the best unified ideal Hume corpus, Kitcher does not thereby block the possibility that there is no unique best unified explanatory order. If there is more than one way to best unify the ideal Hume corpus, then there will simply be no fact of the matter as to what explains what. This would really make the whole Humean enterprise shaky. It would either lead to *relativism*, or else require that it is a priori true that there can be only one best way to deductively unify our knowledge of the world. Both conclusions are equally unpalatable.

There is, however, a second option available to Humeans: to *deny* the mind-dependence of the explanatory order, *without* conceding that there is an independent causal order. To see how this is possible, we need to go back to Ramsey. As I have already stressed in the final paragraphs of section 5.6.1, what he observed was that if Humeans are to offer an objective account of laws, some substantive metaphysical assumptions should be in place. It is really worth repeating the suggestion he made while discussing the idea of a "best deductive system". He noted:

what is asserted is simply something about the whole world, namely that the true general propositions are of such forms

that they form a system of the required sort with the given proposition in the required place; it is facts that form the system in virtue of internal relations, not people's beliefs in them in virtue of spatiotemporal ones. (1928: 132)

The substantive metaphysical assumption that Humeans need to take on board is that the world has an objective structure, in which (fully mind-independent) regularities stand in certain relations to each other. Ramsey's suggestion offers the *external standard* that is required for an objective account of the explanatory order imposed by the best unified ideal Hume corpus: the best unified ideal Hume corpus is the one that respects the objective nomological structure of the world. All this is no less metaphysical than Kitcher's idealism-prone attitude. But if my analysis has been right, Humeans have to engage in metaphysics. They have just two options: one is, let's say, *realist*, while the other is Kantian-idealist. So what could we gain by a realist metaphysics? If it is accepted that the world has a unified structure of regularities, it makes sense to leave the possibility open that even a best unified ideal Hume corpus might be false: it might not capture this objective structure. Relations of explanatory dependence will still be what Humeans take them to be. Causal dependencies will still mirror explanatory dependencies in the ideal unified deductive system. There will still be an open possibility that there is not just one way to best unify the Humean corpus. But something definite has been gained: explanatory relations are subjected to some external – and mind-independent – standard of correctness: *the nomological structure of the world*.

Notes

Introduction

1. Russell (1918: 180). For a criticism of Russell's views, see Lipkind (1979) and Kline (1985). Russell revised his views in his *Human Knowledge* (1948).
2. Mellor calls these platitudes "the connotations of causation". Menzies takes them to analyse the *concept* of causation, whereas Armstrong thinks that they fix its *reference*. Besides, not all philosophers agree that these are genuine platitudes. These issues are orthogonal to the use I make of these platitudes, and hence I won't discuss them.
3. The idea that causes are "recipes" goes back to Gasking (1955).
4. Menzies (1996) takes the intrinsic-relation intuition to be a platitude of causation. But it is too controversial to be a platitude. Armstrong (1999) stresses the role of regularity among the platitudes of causation.

Chapter 1: Hume on causation

1. *A Treatise of Human Nature* will be designated "T". *An Enquiry Concerning Human Understanding* will be designated "E". *An Abstract to A Treatise of Human Nature* will be designated "A". All page references will be to pages in the Selby-Bigge editions of Hume's works.
2. For a learned account of the theories of causation before Hume, as well as for an account of Hume's reactions to them, see Clatterbaugh (1999).
3. For more on Hume's distinction between "philosophical" and "natural" relations, see Stroud (1977: 89) and Robinson (1962).
4. Hume analyses conceivability in terms of the distinctness of ideas. Since "all distinct ideas are separable from each other" (T: 79), we can conceive the one without the other (or, we can conceive the one with the negation of the other). In particular, since the idea of a cause is distinct from the idea of an effect, we can conceive the one without the other. However, as Stroud (1977: 48–50) argues, Hume seems to run in a circle here. For he starts with a conception of distinct ideas in order to show that being distinct, the idea of the cause and the idea of the effect make it *conceivable* that the cause might not be followed by its usual effect. But what makes two ideas distinct other than that the one can be *conceived* without the other? So Hume does not seem to have an independent criterion of distinctness of ideas, which can then be used to found claims of

- conceivability and inconceivability.
5. Fogelin (1985: 46) has aptly called Hume's strategy at this juncture "*the no-argument argument*".
 6. Hume stresses that an appeal to "powers" is bound to fail for three reasons (cf. T: 91). (a) A "power" is not a "sensible quality" of a thing, like being red or square. Why then should we suppose that it exists? (b) Even if we suppose that it exists, it can only be detected by the presence of other sensible qualities of a thing. But why, Hume asks, do we have to suppose that whenever certain sensible qualities are present, the power is also present? (c) Why should we assume that all occurrences of the same class of sensible qualities associated with a thing are accompanied by the instantiation of one and the same power? Hume's objections are all motivated by the same theme: the existence of powers in objects is not a demonstrative truth. And if it is meant to be justified on the basis of experience, then we run against the problem of circularity that we had with the Principle of Uniformity of Nature. Hume's critique of powers is repeated in his first *Enquiry* section IV, part II.
 7. These ideas have been influenced by Stroud (1977: 59–63). He (1977: 60) rightly takes the demand of second-order justification of the rationality of a mode of inference to be *constitutive* of the traditional conception of reason. Demonstration and intuition were taken, by Descartes and almost anyone else, to be immediately justified by virtue of the fact that their second-order justification was transparent to the mind. Presumably, that demonstration is a rational method of inference was seen immediately by reflection. But even if all this is granted, causal inference fails to meet this criterion of second-order justification.
 8. Stroud (1977: 64–5) rightly dismisses the view that principle (R) is an analytic truth.
 9. This objection to Hume has been forcefully made by Stove (1965). Stove argues that Hume offers no argument against what Stove calls "Inductive Probabilism" (IP); namely, the thesis that there are probable inductive arguments. Stove claims that "Hume's refutation of IP is an entirely imaginary episode in the history of philosophy" ([1965] 1968: 189). For a critique of Stove's account of Hume, see Fogelin (1985: 154–7).
 10. For a relevant discussion see Gillies (2000: Ch. 3).
 11. We could, as Carnap (1950) in effect does, put a premium on *E* as opposed to any of the *E_i*s, based on the fact that, in the past, *C*s have been associated with *E*s. But what can justify this premium if not the claim that the future is more likely to be the same as the past than not?
 12. Hume makes no space for synthetic a priori principles, that is, of principles that are synthetic (i.e. refer to matters of fact) and are ascertainable (knowable) a priori. In a famous passage of the *Enquiry* (E: 25) he says: "All the objects of human reason or enquiry may naturally be divided into two kinds, to wit, *Relations of Ideas* and *Matters of Fact*". Relations of ideas "are discoverable by the mere operation of thought, without dependence on what is anywhere existent in the universe" (*ibid.*). Matters of fact "are not ascertained in the same manner; nor is our evidence of their truth, however great, of a like nature with the foregoing [relations of ideas]. The contrary of every matter of fact is still possible; because it can never imply a contradiction, and is conceived by the mind with the same facility and distinctness, as if ever so conformable to reality" (E: 25–6). So relations of ideas belong to the realm of reason and are knowable a priori, whereas matters of fact belong to the realm of experience and are knowable a

posteriori. This bifurcation leaves no space for a third category of synthetic a priori principles. Immanuel Kant (1787, B5, A9/B13, B124) did, famously, make space for synthetic a priori judgements and, unlike Hume, thought that causal inference does constitutively involve some kind of necessity which is like necessity₂, but synthetic a priori in character. There is a massive literature on Kant's views on causation. An important recent piece that discusses the relation between Hume and Kant is Falkenstein (1998).

13. A critical account of Mackie's reading of Hume is given by Beauchamp & Rosenberg (1977: 373–9).
14. This aspect of Hume's account was first stressed by Kemp Smith in his monumental work on Hume (1941). Indeed, Kemp Smith was the first philosopher to bring to focus this naturalistic element of Hume's philosophy. He notes: "This observation of repeated sequence generates – *causally* generates – in the mind a custom or habit. This custom or habit, in turn, itself generates – again in a *causal* manner – the feeling of *necessitating* transition; and it is upon the pattern of this impression that our ideas of causal connexion have come to be modeled" (1941: 373).
Kemp Smith's approach has also been adopted by Stroud (1977: 92) and by Fogelin (1985: 48–9).
15. Stroud (1977: 92–3) goes on to argue why Hume's critique of causation does not create problems for his own *causal* theory.
16. Has Hume here confused two distinct senses of "foundation", the first being logical and the second psychological? Not really. As Craig (1987: 85) has rightly observed, Hume here states one of "the most basic points of his philosophy"; namely, that "where philosophers thought that they saw the operations of reason, the divine spark at work in man, they were watching nothing more than a mundane mechanism and its natural effects in the mind".
17. The first definition runs thus: "we may define a cause to be *an object, followed by another, and where all the objects similar to the first are followed by objects similar to the second*" (E: 76). Note that the requirement of spatial contiguity has faded away from this definition, although the requirement of temporal succession has remained. As for the second definition, it goes as follows: "*an object, followed by another, and whose appearance always conveys the thought to that other*" (*ibid.*). Note that in the wording of the second definition, the reference to "determination" has been dropped.
18. There have been three major interpretative strategies. The *first* is to argue that Hume intended only one of the two definitions as the *proper* definition of causation. The *second* is to try to show that the two definitions are, essentially, equivalent. The *third* is to deny that either of the two is a proper definition.

Among the followers of the first strategy there is a division between those who take Hume to have held only Df_1 as the proper definition of causation and those who take him to have asserted only Df_2 as the proper definition. According to Robinson (1962), Hume was a defender of RVC. He then takes it to be the case that Hume was in "error" when he offered Df_2 as a definition of causation. (Robinson's account is criticized, not altogether successfully, by Richards (1965). See Robinson (1968) for his reply.) Kemp Smith (1941), on the other hand, was among the first who emphasized that Hume was not a defender of RVC. So he is among those who argue that the only proper definition of causation is Df_2 . In defending Df_2 , Kemp Smith (1941: 401) makes the startling suggestion that Df_2 is really an "ostensive" definition. Hume is said to invite us to

see what causation is by looking at the *causal* connections between some events in our imagination.

Among those who take Hume to have offered two strict – and ultimately equivalent – definitions of causation the most notable attempt is by Garrett (1993). Concerning Df_2 , Garrett argues that the reference that it makes to the mind can be construed as a reference to a “generalised or ‘ideal’ mind or spectator” (1993: 180). But Garrett’s key move is to argue that, far from involving no reference to necessity, Df_1 can be seen as a way to “comprehend” the necessary connection between causally related events – even if, for Hume, the idea of necessary connection turned out to be something other than what was expected. For Garrett, where Df_2 “characterises the set of *ideas* that give rise to the internal impression of necessary connection directly”, Df_1 “characterises the same set indirectly, by characterising the *objects* whose ideas give rise to this impression” (1993: 184). So both Df_1 and Df_2 are taken to be exact definitions of the very same notion, and in particular of the concept of causation as involving the idea of necessity.

- Finally, Stroud (1977: 89) argues that neither of the two definitions should be seen as a proper definition in the sense of defining the meaning of ‘*C causes E*’.
19. Stroud (1977: 90–91) rightly stresses that Hume leaves entirely open the possibility that there might be beings whose minds are not constituted like ours, and which therefore lack the idea of necessary connection.
 20. A similar view is suggested by Beauchamp & Rosenberg (1981: 28–31).
 21. For a telling criticism of Hume’s theory of ideas, see Stroud (1977: 229–30).
 22. A similar thought is expressed by Broughton (1987). She, however, stresses that all we can attribute to Hume is that we can achieve “the *bare thought* . . . of there being some feature of objects that underlies . . . constant conjunction” (1987: 126).
 23. Craig (2000) argues that Hume might be seen as being *both* a realist and a projectivist.
 24. Important articles on the New Hume debate can be found in Read and Richman (2000).

Chapter 2: Regularities and singular causation

1. The word *factor* is used as a catch-all term to cover causal antecedents. Depending on one’s philosophical preferences, one can substitute either *property* or *event-type* for *factor*.
2. Mill (1911: 218) does try to explain how this claim about counteracting *causes* does not affect his account of causation.
3. If unconditionality fails, Mill (1911: 253) notes, then “both the antecedent and the consequent might be successive stages of the effect of an ulterior cause”.
4. Unlike Mill, Ducasse (1969: 20) thinks it *does* make sense to distinguish between *the* cause of an effect and its conditions.
5. While Ducasse accepts that causal laws are regularities, it is possible for an advocate of singular causation to take causal laws to be of a different sort. Indeed, David Armstrong (1997) does accept the existence of singular causation and yet takes causal laws to be relations among universals. (His views are examined in some detail in Chapter 6.)
6. For a more sympathetic – yet also critical – reading of Ducasse’s views, see Mackie (1974: 135–42).
7. This claim can be contested. There has been some interesting experimental work

(done mostly by Albert Michotte after the Second World War) that suggests that human beings have, after all, a direct impression of the influence of one physical object on another. As Michotte put it: “certain physical events give an immediate causal impression, and . . . one can ‘see’ an object *act* on another object, *produce* in it certain changes, and *modify* it in one way or another” (quoted by Hewstone 1989: 6). His critics, however, point out that these results can be the product of projecting anthropomorphic concepts onto events in the world.

8. Stroud (1977: 230–1) goes as far as to argue that if we lacked the idea of necessary connection, we would also lack “all such [causal] verbs”.
9. Her favourite example is the concept of work. *Work* is an abstract concept whose manifestations are writing papers, doing the dishes, doing DIY repairs, going to meetings, teaching and so on. The abstract/concrete relation is further elaborated in Cartwright (1999: 37–43).
10. For a criticism of the view that causal relata are facts, see Davidson ([1967] 1993: 77–8) and Hausman (1998: 22–3). Davidson’s main point is that if we took the relata of causation to be facts, the extensional character of causal statements would be lost. Mellor responds to Davidson’s criticism in Mellor (1995: Chs 10 & 11).
11. For an important recent defence of singular causation, see Cartwright (2000a).

Chapter 3: Causation and counterfactuals

1. For more on this see Beauchamp & Rosenberg (1977: 400ff).
2. Armstrong (1983: 51) raises a similar point, but he phrases his objection in a different way. He claims that on Mackie’s account, the relevant counterfactuals would have *probabilistic consequents*, that is they would be of the form: “if *x* had been an *F*, then it is *very probable* that *x* would also have been a *G*”. But, as David Papineau has noted to me, these sort of counterfactuals are just what is needed for causal sequences according to standard counterfactual theories of probabilistic causation.
3. The term “fragile” is from Lewis (1986c: 198). There are also other reasons for rejecting extremely fragile events, for example, that they may give rise to spurious causal dependencies. Here is Lewis’s memorable example: “Boddie eats a big dinner, and then the poisoned chocolates. Poison taken on a full stomach passes more slowly into the blood, which slightly affects the time and manner of the death. If the death is extremely fragile, then one of its causes is the eating of the dinner. Not so” (*ibid.*).
4. It would be wrong to think that Mackie’s need for an account of the direction of causation arises only in connection with his attempt to deal with overdetermination. It is also required by his attempt to link counterfactuals with regularities. Why, for instance, shouldn’t we assert the counterfactual *if the match had been struck, it wouldn’t have been dry*? This seems to be supported by a perfectly good law, which one might carry over to other possible worlds; namely, that matches that (a) are struck and (b) don’t light, and (c) have oxygen around, must (d) be wet. (a) is the counterfactual supposition and (b) and (c) are other features of the actual world which are *held fixed* in going to other (a)-worlds. Mackie should then claim that one mustn’t hold fixed things like (b) which are *effects* of the antecedent (a). But this assumes some hold on causal direction: so the regularities that support counterfactuals had better be directed. (I owe this point to David Papineau).
5. For more on causal overdetermination, see Scriven (1966).

6. For more on this see Mackie (1974: 64–71).
7. Lewis (1986c: 167) considers the idea of “nomic dependence” and tries to see how far it can go in capturing causal dependence. He argues that although causal dependence is irreversible, nomic dependence is not. For a criticism of Lewis’s views and a defence of nomic dependence, see Beauchamp & Rosenberg (1981: 164–8), Clendinnen (1992) and Kim (1973: 571–2).
8. Lewis’s criteria for the similarity among possible worlds are, partly, introduced to account for the asymmetry of counterfactual dependence. For a criticism see Horwich (1987: 172–3).
9. Lewis’s (1986c) initial attempt involved the thought that there is a special kind of dependence between the pre-empting cause *c* and the effect *e*, which he called “quasi-dependence”. This kind of dependence is causal by courtesy: because it resembles cases of genuine causal dependence. For a repudiation of this view, see Lewis (2000: 184–5). Other attempts include Menzies (1989) and McDermott (1995). For a general criticism of all these attempts see Schaffer (2000).
10. It should be noted, however, that as Ehrling (1997: Ch. 3) argues, pre-emption may well be a problem for all theories of causation.
11. For further discussion of these issues, as well as for learned accounts of Lewis’s approach, see Owens (1992: 49–60) and Hausman (1998: Ch. 6).
12. It should be noted that Menzies and Price’s account is stated in probabilistic terms in order to accommodate indeterministic causation.
13. However, Menzies & Price (1993) engage in a systematic attempt to dispel one by one all of the objections levelled against the agency theory.
14. Thanks to my colleague Panagiotis Oulis for drawing my attention to this example.
15. In a randomized experiment, the subjects are *randomly* selected and *randomly* divided into two groups, the experimental group and the control group. Suppose one wants to test the effects of a new pill on the recovery from a certain disease. The experimental group is given, say, the pill whose effects need to be tested, whereas the control group is given a placebo (e.g. a sugar pill). (The experiment is double-blind because the subjects do not know whether they are given the pill or the placebo and the experimenters do not know whether they give the pill or the placebo.) This experimental design makes it very likely that the *placebo*-effect (e.g. getting better just because of the self-inflicted expectation of getting better because a pill was administered) will be the same in both groups, as will be any other variables, known as *confounding* variables, which might be correlated with recovery. This last thing is ensured by the random assignment of individuals to the two groups. Then, if there is a difference (actually, a statistically significant difference) in recovery among those who took the pill and those who took the placebo, this difference will be very likely owing to the pharmacological effect of the pill. This kind of experiment ensures *modularity* by making sure that the pharmacological mechanism by which the pill leads to recovery is present only in the experimental group and absent in the control group. For useful introductions into the many aspects (and problems) of experimental design, see King *et al.* (1994) and McGuigan (1997).
16. Cartwright (2000b) offers a detailed critique of Hausman’s and Woodward’s interventionist account.

Chapter 4: Causation and mechanism

1. For a similar criticism see Beauchamp and Rosenberg (1977: 398).
2. A notable recent attempt to improve on Mackie's ideas is Ehrling's (1997). He takes over Mackie's idea of causation-as-persistence and argues that what persists in a causal sequence is some *property*. As he puts it: a causal process is "a process that extends from point *A* to point *B* such that there is some property *P* which characterises this process at every point between and including *A* and *B*" (1997: 120). To be sure, Ehrling refines this claim by taking the causal relata to be *tropes* and not properties, where tropes can be usefully thought of as particularized properties. Like universals, tropes can persist over time. But unlike universals, tropes cannot characterize more than one individual at the same time.
3. An earlier version of the *process theory* of causation was offered by Russell (1948). His views are analysed and criticized by Salmon (1984: 144–6) and Dowe (2000: 62–6).
4. In an earlier piece, Salmon noted that "the mark method may be said, roughly speaking, to provide a means for distinguishing causal regularities from other types of regularity in the world, including those that may be associated with pseudo-processes" (1997a: 198).
5. The discussion of the relevant counter-examples is given in Salmon (1984: 171–4).
6. Given the modification of Salmon's theory, mark-transmission (MT) should also be modified. Salmon characterizes MT in terms of intersections, but now it should be recast in terms of intersections.
7. An important aspect of Salmon's theory, which we shall not discuss, concerns the "production" of causal processes. Salmon's main idea is that the "production of structure and order" in the world is, at least partly, due to the existence of "conjunctive forks", which are exemplified in situations in which a common cause gives rise to two or more effects. The core of this idea goes back to Reichenbach (1956), though Salmon also adds further cases of causal forks, such as "interactive forks" and "perfect forks", which correspond to different cases of common-cause situations. Salmon uses statistical relations among events to characterize causal forks. He also argues that it is the *de facto* direction of the causal forks from past events to future events that constitutes the direction of causation. For the details of these views, see Salmon (1984: Ch. 6, 1997a: Ch. 18). For criticisms, see Dowe (2000: 79–87).
8. Salmon (1984: 149–50, 1997a: 253) makes an effort to offer a quasi-objective account of the truth-conditions of counterfactuals. For an apposite criticism, see Dowe (2000: 84–5).
9. Before adopting a version of the Conserved Quantity theory, Salmon (1997a: Ch. 16) advanced the view that causation involves invariant (and not just conserved) quantities. But in light of Dowe's (2000) and Hitchcock's (1995) criticisms, he abandoned it in favour of the Conserved Quantity theory. For an exposition and criticism of the Invariant Quantity theory, see Dowe (2000: 114–19). Dowe defends his own version of the Conserved Quantity theory against Salmon's, as well as against the earlier *transference* and *persistence* accounts, in his *Physical Causation* (2000: 109–22).
10. Some recent work on mechanistic accounts of causation includes Glennan (1996) and Machamer *et al.* (2000).
11. There are some complications. Lewis (1983) has made a distinction between

relations that are intrinsic to their *relata* and relations that are intrinsic to their *pairs*. A relation R is intrinsic to its *relata* a and b if the following holds: if there were two things a' and b' such that they are exact duplicates of a and b , then R would also hold for a' and b' . A relation R is intrinsic to its *pairs* if the following holds: if R holds for the pair $\langle a, b \rangle$, then if there was an exact duplicate pair $\langle a', b' \rangle$, R would hold for this too. A relation can be intrinsic to its *relata* without being intrinsic to its *pairs*. For instance, it may happen that a and b are the exact duplicates of a' and b' but the spatiotemporal distance between a and b is different from the spatiotemporal distance between a' and b' . In this case, $\langle a, b \rangle$ and $\langle a', b' \rangle$ would not be duplicate pairs. All this is relevant because the advocates of the claim that causation is an intrinsic relation (e.g. Menzies 1999: 320) argue that causation is intrinsic to its *pairs* but not to its *relata*. That is, events c and e might be the exact duplicates of events c' and e' , but it may be that although c causes e , c' does not cause e' . (For instance, e' might occur earlier than c'). Practically, this means that the *properties* of the cause and the effect are not enough to make causation an intrinsic relation. The *relations* (spatiotemporal and others) between the cause and the effect are also required to make causation an intrinsic relation.

12. To be sure, Menzies & Price (1993: 197) suggest a way in which their agency theory might offer an *intrinsic* characterization of causation. Their thought is that the manipulability of causes is *supported* by the intrinsic features of the causal relation between events.
13. Menzies (1999) tries to offer a detailed account of what exactly it is for causation to be an intrinsic relation. But his views depend on taking as primitive that there are natural properties of, and *natural relations* among, the events that are causally related and that causation *supervenes* on them.
14. Some cogent arguments against Tooley's "causal realism" are offered by Ehring (1997: 66–8).
15. There is a rather important reason why counterfactual theories cannot offer an intrinsic characterization of causation. As it was stressed in section 3.3.2, if causation reduces to counterfactual dependence among events, then the truth of the claim that c causes e will depend on the absence of overdeterminers, since if the effect e is causally overdetermined, it won't be counterfactually dependent on any of its causes. But the presence or absence of overdeterminers is certainly *not* an intrinsic feature of a causal sequence (cf. Lewis 2000: 185).

Chapter 5: The regularity view of laws

1. Beauchamp & Rosenberg (1981: 140ff) try to show how the basics of this view can be found in Hume's work.
2. It is noteworthy that an appeal to counterfactuals is consistent with the epistemic approach discussed in the previous section. In fact, defending the view that laws of nature are inductively established regularities, Mackie (1966) suggested that this view could explain why laws – but not accidents – can support counterfactuals. Suppose, he says, that it has been inductively established that *All Fs are Gs*. If we were then to ask of an individual x , which is not an instance of F , the question "if x were an F , then would x also be a G ?", Mackie says that it would be very natural to think that the answer to this question would be positive. So, he concludes, the inductively established law *All Fs are Gs* supports the counterfactual "if x were an F , then x would also be a G ". However, Mackie's suggestion fails to show that accidents do not support coun-

terfactuals. If, after all, some accidentally true generalizations can be established inductively, then it seems that they will also support the relevant counterfactuals. Further problems with Mackie's views of counterfactuals have been discussed in section 3.1.1.

3. For more discussion of Goodman's theory, see Horwich (1987: 157–8).
4. For a similar criticism, one can look at Salmon (1977: 199–200).
5. It should be noted that Carnap also took laws of nature to be whatever lawlike statements are *deducible* from a set of axioms that express a certain physical theory, or more generally “the deductive system of physics” (1928: 166).
6. The MRL approach is sympathetically discussed in Earman (1984). Incisive criticisms are found in Armstrong (1983) and Menzies (1993).
7. Lewis discusses further his account of laws of nature in Lewis (1986e: 122–4).
8. In his (1929: 138), Ramsey rejected the *best systems* approach he had put forward just a year before. He noted that this view was untenable because “it is impossible to know everything and organise it in a deductive system”. In its place, he put the view that laws of nature are distinguished from accidents on the basis that we trust the former but not the latter. He therefore noted that his later view was closer to the epistemic approach. However, in an interesting paper, L. J. Cohen (1980) has suggested that Ramsey's later view is much closer to his earlier one than it first appears.
9. Carroll (1990) offers an incisive criticism of Lewis's approach to laws.
10. The role of natural properties in Lewis's approach to laws is discussed in detail by Carroll (1990) and Loewer (1996). Carroll doubts that Lewis's account of natural properties holds much promise, while Loewer offers an account of natural properties that can help Lewis.
11. Tooley (1977) has insisted that laws cannot be MRL-regularities because he thinks that there are (can be) uninstantiated laws, which cannot be derived within the best system. For a telling criticism of Tooley's views see Earman (1984: 208–10) and Mellor (1991: 149–51).

Chapter 6: Laws as relations among universals

1. Kneale (1949: 79–80) tried to argue against the Humean view that lawlike statements cannot be necessary because their negation is conceivable. His main point was that lack of conceivability has nothing to do with necessity. Take Goldbach's conjecture that every even prime number greater than 2 is the sum of two primes. Given that this is a mathematical statement, and that it is either true or false, if it is true it is necessarily true, and if it is false it is necessarily false. However, Kneale argues, we can conceive the falsity of Goldbach's conjecture, which, if true, is necessarily so. In a similar fashion, he thought, from the fact that we can conceive the falsity of a lawlike statement, it does not follow that if it is true it is not necessarily true. This argument led philosophers such as Ayer (1972: 15) and Popper (1959: 428–9) to conclude that Kneale thought that laws of nature were logically necessary.
2. An interesting, if somewhat convoluted, attempt to strike a balance between Kneale and his Humean opponents is offered by Korner (1953).
3. John Carroll (1990, 1994) disagrees. He takes the ADT approach to be reductive, too.
4. Defending the MRL view against Tooley's example, Earman (1984: 209–10) questions Tooley's assumption that the X-Y law is indeed underived. For more criticism of Tooley's example, see Smart (1993).

5. The *inference problem* is an old one. It first arose in connection with the medieval discussions about universals. A first attempt at a solution is known as the *dictum de omni*, which in the *Port-Royal Logic* appears as follows: "whatever applies to an idea taken universally also applies to everything of which this idea is affirmed, or which is the subject of this idea, or which is included in the extension of this idea" (Arnault & Nicole [1683] 1996: 147). This dictum was taken to be a metaphysical truth. It was severely criticized by Mill (1911: 114–15), who argued that if it is seen as a statement about classes and their members, it is trivially true, but if instead it is seen as a principle expressing "the intercommunity of nature" (that is, the connection between objectively existing universals and particulars classed under them), it is deeply problematic. For more on this issue, see Wilson (1999: 229–36).
6. The same problem has been raised by Lewis (1983: 366). He says: "Whatever N may be, I cannot see how it could be absolutely impossible to have $N(F, G)$ and Fa without Ga ".
7. A nice overview of the problem of universals is given in Armstrong (1989); see also Oliver (1996) and Bealer (1998).
8. This principle is by no means uncontroversial, especially when it comes to universals. Quine, for instance, resists the thesis that since we can make true statements that involve predicates, we should be ontologically committed to the existence of properties as self-subsisting universals.
9. A vigorous defence of the claim that causal relations are perceptually given to us and that we, at least occasionally, are directly aware of them, is given by Fales (1990: Ch. 1). He takes the feeling of pressure on one's own body to be the paradigm case of a perceptually given causal relation. More generally, he (1990: 48) takes forces to be a genus of causal relations, "whose species are forces of specific magnitude".
10. Earman (1984: 217) rightly objects that even if we were to grant that laws involve relations among universals, the ultimate laws of nature will most likely make reference to unobservable entities, and hence the relevant universals will have instances that correspond to not directly observable states of affairs. Hence, their knowledge can only be inferential. Incisive criticism of Armstrong's appeal to direct perception of singular causings is offered by Menzies (1993: 202–3).
11. Here Armstrong disagrees with Tooley (1977, 1987), who admits the existence of uninstantiated universals.
12. To be sure, Armstrong (1997) does try to deal with this issue. But the interested reader had better judge for herself how successful his attempt is.
13. It is worth noting that Dretske (1985) questions Armstrong's claim that there is a necessitating relation between properties, which is, nonetheless, *contingent*.
14. While Armstrong takes the relation of nomic necessitation to be what distinguishes laws from accidents, Tooley (1977: 679) thinks that nomic necessitation is just one among the many kinds of nomological relations that hold among universals and are capable of providing truth conditions to law statements. He is forced to say this because there are apparent differences between several types of law statement. Some have the form "All As are Bs ", others have the form "All As are *not*- Bs ", while others might have the form "All As are *either* Bs or Cs ". If all of these types of statements were taken to have relations among universals as their truth-makers, then it seems that we would have to adopt the unpalatable view that there are negative and disjunctive universals. In order to avoid this metaphysical commitment, Tooley suggests

that there are several types of nomological relations among universals, one being nomic necessitation (giving rise to the regularity of the form “All As are B”) another being nomic exclusion (giving rise to the regularity of the form “All As are not-B”) and so on. Despite the ingenuity of Tooley’s technical account of the several types of nomic connections among universals, one cannot avoid the feeling that all this is a rather *ad hoc* way to capture the intuitive pull of the idea that there must be something that makes a law a law and not an accident.

15. For a discussion and defence of metaphysical necessity, see Lowe (1998: 8–21) and Ellis (2000: 335–7).
16. For a recent argument that at least some laws are metaphysically necessary, see Bird (2001). Beebe (2002) and Psillos (2002b) reply to Bird’s argument, and Bird (2002) defends further his position.
17. Mumford (1998: 233–6) offers a limited defence of dispositional essentialism. He also takes the view that laws of nature can be seen to derive from the basic dispositions of things.
18. For a recent challenge of the presuppositions of Carroll’s counter-example, see Beebe (2000).

Chapter 7: Alternative approaches to laws

1. For more on the distinction between enumerative and eliminative induction (cast within the setting of inference to the best explanation) see Psillos (2002a).
2. See Carnap (1937: 321) for a discussion of Schlick’s position.
3. For a criticism of the inference-ticket view, see Hempel (1965: 354–9).
4. There is a *third* characteristic too; namely, that the intervention *I* is not correlated with other causes of *Y* besides *X*.
5. Woodward’s views about the connection between active counterfactuals and the distinction between laws and accidents are, in fact, more complicated, and explained in Woodward (2000: 237–8).
6. In fact, Woodward (2000: 206–7) too argues that this law cannot be accounted for in terms of invariance.
7. The view that laws are invariances of some sort has also been defended by Skyrms (1980). He frames the connection in terms of *resiliency*, this being the property that the epistemic probability of a statement has to remain stable in several contexts. Woodward (1992: 215, n.11) disagrees with the Skyrmsian account because it relies on epistemic notions. For a criticism of Skyrms’s account, see Carroll (1990: 201–11).
8. For an important survey of the debate around *ceteris paribus* laws, as well as a defence of strict laws in physics, see Earman & Roberts (1999).
9. All these theses are defended in Cartwright (1989: 1–9), but versions of them are also in Cartwright (1999).
10. For Cartwright’s reaction to this, see Cartwright (1995b).
11. Sometimes, talk of dispositional properties is contrasted with talk of occurrent properties, that is, of properties that are *actually* possessed by an object. Yet, as C. B. Martin (1994) has pointed out, the defenders of dispositions take dispositions to be actually possessed by the objects, even if they may not be, at a time, manifested.
12. This view is defended by Harré (1970).
13. For a passionate defence of irreducible dispositions, see Mellor (1991: Ch. 6).
14. Lange, to be sure, doubts that laws must be seen as exceptionless regularities. Although he leaves it open that there might be such regularities, he argues that

belief in exceptionless regularities is not part of the root commitment undertaken when it is believed that a law obtains (cf. 2000: 161). Laws can be violated but, as he stresses, this is not a problem for their special function in science, if (a) these violations are “offstage”, that is, they are “of no concern to us” (2000: 27); and (b) if the law is generally reliable.

15. In an earlier piece though, Lange (1993) did seek to defend a modern version of the inference-ticket view of laws. His disagreement with the traditional inference-tickets view is explained in Lange (2000: 188–90).
16. However, Lange (2000: 108–9) takes this uniqueness claim back slightly. That is, he allows that there may be possible worlds with “multiple grades of necessity” between conceptual necessity and no necessity at all.
17. Indeed, Lange’s (2000: 106) account of stability is based on the assessment of *nested* counterfactuals of the form “if *p* had been the case, then if *m* had been the case then *L* would still be a law”. It is obvious that the assessment of such counterfactuals (and of even more complex ones) is precarious without an account of what makes them correct.
18. Mellor’s account of universals (1991: 152–3) is modelled on Ramsey’s (1925). This is interestingly different from Armstrong’s account, although it also takes it to be the case that there are no uninstantiated universals.
19. In fact, Mellor (1991: 163) takes it that the law asserts something stronger, namely the counterfactual that for *anything*, whether it is *F* or not, if it were *F* it would be *G*.
20. For more on this see Psillos (1999: Ch. 3).
21. The interested reader should also see Mellor (1995: Chs 15 & 16).
22. Van Fraassen disagrees. He argues (1989: 183) that “there are no laws of nature” and offers a sustained argument to the effect that the concept of law is metaphysically problematic and that it is neither necessary to, nor really useful for, science. A critique of van Fraassen’s views is offered by Earman (1993) with a response in van Fraassen (1993).

Chapter 8: Deductive–nomological explanation

1. The recent articles by Sklar (1999) and Kim (1999) cast further light on this issue.
2. Davidson (1967) famously drew a dichotomy between causation and causal explanation. The former, he suggested, relates *events*, whereas the latter relates *facts*. For a thorough criticism of Davidson’s view, see Mellor (1995: 130–31). As Mellor stresses, (a) causes explain their effects, and (b) facts can be causes. So there is an intimate connection between causation and causal explanation, and consequently, Davidson’s dichotomy is far from compelling. For an attempt to strike a balance between Davidson’s views and Mellor’s, see Hausman (1998: 22–3).
3. Elsewhere (1965: 362), Hempel noted that “explanatory statements of the form ‘*q* because *p*’” should be construed as follows: “*p* is (or was) the case, and there are laws (not explicitly specified) such that the statement that *q* is (or was) the case follows logically from those laws taken in conjunction with the statement of *p* and perhaps other statements, which specify antecedents not included in *p* but tacitly presupposed in the explanation”.
4. Although the *flagpole* version became famous, Bromberger’s original story concerned the height and the shadow of the Empire State Building.
5. Mathematical explanation is a clear case of non-causal explanation; as is the

- case in which one explains why an event happened by appealing to conservation laws, or to general non-causal principles (such as Pauli's exclusion principle).
6. Van Fraassen (1980: 132–4) disagrees. He tells the following story about the explanatory relation between the length of the shadow and the height of a tower. A knight killed the maid with whom he was in love. The tower he built subsequently marks the spot where he killed her. Why does the tower have the height it does? In order for it to cast a shadow of a certain length at noon, so that, every noon, the shadow covers the terrace where the knight first proclaimed his love to the maid. In this case, it is the length of the shadow that explains the height of the tower. Van Fraassen's pragmatic theory of explanation is criticized by Kitcher & Salmon (1987).
 7. For a brief but incisive account of "where our idea of a time asymmetry in explanation comes from", see Sklar (1999: 31–2). For a defence of a neo-Humean theory of the direction of causation, see Horwich (1987: 132–8). For a critique of Horwich's position, see Hausman (1998: 44–7).
 8. For a thorough defence of this point, as well as of the DN model, see Redhead (1990).
 9. An important aspect of Scriven's critique of the DN model – which later on reappeared in Cartwright (1983) – is that the laws quoted in a DN explanation are not "literally true" (1962: 100).
 10. Even if one were to take the currently popular view that manipulation requires only *invariant relations* among magnitudes or variables, and even if it was admitted that these invariant relations do not hold universally but only for a certain range of interventions/manipulations, one would still be short of a genuinely singularist account of causation.
 11. Woodward (1984) tried to reject the "conventional view" that "singular causal explanations are in some way truncated or implicit covering-law explanations". In its place, he puts the view that singular causal explanations are "a distinct genre of explanation, which does not possess anything remotely like a covering law structure" (1984: 247). So Woodward thinks that claims such as *the short circuit caused the fire* offer complete explanations of the relevant singular events. For a telling criticism of Woodward's view, see Kim (1999: 15–17). Woodward defends his position further in Woodward (1986).
 12. Another type of counter-example comes from the existence of cases that provide a necessary but not sufficient condition for the occurrence of an event. Among the people who suffer from latent untreated syphilis, there is a 25 per cent probability that they will develop general paresis. So, the hypothesis that a subject *S* has latent untreated syphilis cannot be used to predict that the subject will develop general paresis. After all, only one in four do. Yet having latent untreated syphilis is a necessary condition for developing general paresis, which means that after a subject has been diagnosed with general paresis, we can infer that the subject had syphilis. So we can explain why the subject has general paresis by saying that the subject has latent untreated syphilis, although we cannot predict the former from the latter. This counter-example is criticized by D. B. Hausman (1976).

Chapter 9: Statistical explanation

1. I am saying that he *toyed* with the idea, as opposed to explicitly endorsing it, because although he cited von Mises' endorsement of a statistical understanding of causation, he didn't seem willing fully to endorse his view.

2. An important recent book on the philosophical theories of probability is Gillies (2000).
3. The exact definition, which is slightly more complicated and accurate than the one offered here, is given by Hempel (1965: 400). An excellent detailed account of RMS is given in Salmon (1989: 55–7).
4. The concept of objective homogeneity and its implications are discussed in Salmon (1984: Ch. 3).
5. It might be thought that only relations of positive statistical relevance are explanatory. Yet Salmon (1984: 46) suggests that negative relevance is also explanatory. To use Humphreys' (1981) felicitous terminology, relations of positive relevance can express "contributing causes", while relations of negative relevance express "counteracting causes".
6. The *Simpson paradox* suggests that *C* may cause *E*, even though *C* is not statistically correlated with *E* in the whole population. For more on this see Cartwright (1983: essay 1) and Suppes (1984: 55–7).
7. Not all agree with this view. For the defence of a kind of reductive view about probabilistic causation, see Papineau (1989).
8. For further discussion of this issue, see Salmon (1989: 154–66).
9. For a recent interesting attempt to defend a contrastive account of causal explanation, see Ylikoski (2001).

Chapter 10: Explanation of laws

1. This tradition goes back to Mach and Poincaré, but Friedman wants to dissociate the idea of unification from Mach's and Poincaré's phenomenalist or instrumentalist accounts of knowledge.
2. The basic idea can be found in Feigl (1970: 12). He takes "the fact–postulate ratio" to represent "the explanatory power of theories". As he put it: "The aim of scientific explanation throughout the ages has been *unification*, i.e., the comprehending of a maximum of facts and regularities in terms of a minimum of theoretical concepts and assumptions".
3. Of course, when it comes to natural objects, there are no circumstances under which they were "designed". But, as Kitcher (1989: 485) notes, the situation is different only in its degree of complication.
4. Barnes (1992) has presented some very interesting objections to Kitcher's attempt to show how his theory can solve the problem of the asymmetry of explanation. The essence of these objections is that there are plenty of effect-to-cause argument-patterns that are maximally unifying. Hence, Kitcher cannot easily dismiss them as non-explanatory.

Chapter 11: The metaphysics of explanation

1. The modal conception is defended by Mellor (1995: 75–6), who also expands it so that it covers indeterministic causal explanation. See also Hausman (1998: 160).
2. However, the notion of explanatory dependence is broader than the notion of causal dependence: there may be explanatory dependencies that are not causal (e.g. explanatory dependencies among laws).
3. Kitcher (1986) defends this neo-Kantian line. As he clearly explains (1986: 210), Kantians take laws of nature to be "mind-dependent".

References

- Achinstein, P. 1983. *The Nature of Explanation*. New York: Oxford University Press.
- Anscombe, G. E. M. 1971. "Causality and Determination". See Sosa & Tooley (1993), 88–104.
- Armstrong, D. M. 1978. *Universals and Scientific Realism* [2 volumes]. Cambridge: Cambridge University Press.
- Armstrong, D. M. 1983. *What is a Law of Nature?* Cambridge: Cambridge University Press.
- Armstrong, D. M. 1989. *Universals: An Opinionated Introduction*. Boulder, CO: Westview Press.
- Armstrong, D. M. 1993a. "The Identification Problem and the Inference Problem", *Philosophy and Phenomenological Research* 53: 421–2.
- Armstrong, D. M. 1993b. "Reply to Smart". See Bacon *et al.* (1993), 169–74.
- Armstrong, D. M. 1993c. "Reply to Menzies". See Bacon *et al.* (1993), 225–32.
- Armstrong, D. M. 1997. "Singular Causation and Laws of Nature". In *The Cosmos of Science*, J. Earman & John Norton (eds), 498–511. Pittsburgh, PA: Pittsburgh University Press.
- Armstrong, D. M. 1999. "The Open Door: Counterfactual Versus Singularist Theories of Causation". In *Causation and Laws of Nature*, Howard Sankey (ed.), 175–85. Dordrecht: Kluwer.
- Arnault, A. & P. Nicole 1683. *Logic or the Art of Thinking*, Jill van Buroker (ed.). Cambridge: Cambridge University Press, 1996.
- Ayer, A. J. 1963. "What is a Law of Nature?". In *The Concept of a Person and Other Essays*, A. J. Ayer, 209–34. London: Macmillan.
- Ayer, A. J. 1972. *Probability and Evidence*. London: Macmillan.
- Bacon, J., K. Campbell, L. Reinhardt (eds) 1993. *Ontology, Causality and Mind: Essays in Honour of D. M. Armstrong*. Cambridge: Cambridge University Press.
- Barnes, E. 1992. "Explanatory Unification and the Problem of Asymmetry", *Philosophy of Science* 59: 558–71.
- Bealer, G. 1998. "Universals and Properties". In *Contemporary Readings in the Foundations of Metaphysics*, S. Laurence & C. Macdonald (eds), 131–47. Oxford: Blackwell.
- Beauchamp, T. L. & A. Rosenberg 1977. "Critical Notice of J. L. Mackie's *The Cement of the Universe*", *Canadian Journal of Philosophy* 7: 371–404.

- Beauchamp, T. L. & A. Rosenberg 1981. *Hume and the Problem of Causation*. Oxford: Oxford University Press.
- Beebe, H. 2000. "The Non-Governing Conception of Laws of Nature", *Philosophy and Phenomenological Research* 61: 571–94.
- Beebe, H. 2002. "Contingent Laws Rule", *Analysis* 62: 252–5.
- Bennett, J. 1987. "Event Causation: The Counterfactual Analysis", *Philosophical Perspectives* 1. Reprinted in Sosa & Tooley (1993), 217–23.
- Bird, A. 2001. "Necessarily, Salt Dissolves in Water", *Analysis* 61: 267–74.
- Bird, A. 2002. "On Whether Some Laws are Necessary", *Analysis* 62: 257–70.
- Blackburn, S. 1984. *Spreading the Word*. Oxford: Oxford University Press.
- Blackburn, S. 1990. "Hume and Thick Connexions", *Philosophy and Phenomenological Research* 50 (supplement): 237–50.
- Braithwaite, R. B. 1953. *Scientific Explanation*. Cambridge: Cambridge University Press.
- Bromberger, S. 1966. "Why-Questions". In *Mind and Cosmos: Essays in Contemporary Philosophy of Science*, R. G. Colodny (ed.). Pittsburgh, PA: Pittsburgh University Press.
- Broughton, J. 1987. "Hume's Ideas About Necessary Connection", *Hume Studies* 13: 217–44.
- Butterfield, J. 1985. "Review of *What is a Law of Nature?* by D. M. Armstrong", *Mind* 94: 164–6.
- Carnap, R. 1928. *The Logical Structure of the World*. Berkeley, CA: University of California Press.
- Carnap, R. 1937. *The Logical Syntax of Language*. London: RKP.
- Carnap, R. 1950. *Logical Foundations of Probability*. Chicago, IL: Chicago University Press.
- Carnap, R. 1974. *An Introduction to the Philosophy of Science*, New York: Basic Books.
- Carroll, J. W. 1987. "Ontology and the Laws of Nature", *Australasian Journal of Philosophy* 65: 261–76.
- Carroll, J. W. 1990. "The Humean Tradition", *The Philosophical Review* 99: 185–219.
- Carroll, J. W. 1994. *Laws of Nature*. Cambridge: Cambridge University Press.
- Cartwright, N. 1983. *How the Laws of Physics Lie*. Oxford: Clarendon Press.
- Cartwright, N. 1989. *Nature's Capacities and Their Measurement*. Oxford: Clarendon Press.
- Cartwright, N. 1993. "In Defence of 'This Worldly' Causality: Comments on van Fraassen's *Laws and Symmetry*", *Philosophy and Phenomenological Research* 53: 423–9.
- Cartwright, N. 1995a. "Précis on *Nature's Capacities and Their Measurement*", *Philosophy and Phenomenological Research* 55: 153–6.
- Cartwright, Nancy 1995b. "Reply to Eells, Humphreys and Morrison", *Philosophy and Phenomenological Research* 55: 177–87.
- Cartwright, N. 1999. *The Dappled World*. Cambridge: Cambridge University Press.
- Cartwright, N. 2000a. "An Empiricist Defence of Singular Causes". In *Logic, Cause and Action*, R. Teichmann (ed.), 47–58. Cambridge: Cambridge University Press.
- Cartwright, N. 2000b. *Measuring Causes: Invariance, Modularity and the Causal Markov Condition*. Measurement in Physics and Economics Discussion Papers Series, CPNSS, London School of Economics, DP MEAS 10/00.

- Chappell, V. C. (ed.) 1968. *Hume*. Notre Dame, IN: University of Notre Dame Press.
- Clatterbaugh, K. 1999. *The Causation Debate in Modern Philosophy 1637–1739*. New York: Routledge.
- Clendinnen, F. J. 1992. "Nomic Dependence and Causation", *Philosophy of Science* 59: 341–60.
- Coffa, A. 1974. "Hempel's Ambiguity", *Synthese* 28: 141–63. Reprinted in *Explanation*, D.-H. Ruben (ed.), 56–77. Oxford: Oxford University Press, 1993.
- Cohen, L. J. 1980. "The Problem of Natural Laws". In *Prospects for Pragmatism*, D. H. Mellor (ed.), 211–28. Cambridge: Cambridge University Press.
- Costa, M. J. 1989. "Hume and Causal Realism", *Australasian Journal of Philosophy* 67: 172–90.
- Craig, E. 1987. *The Mind of God and the Works of Man*. Oxford: Oxford University Press.
- Craig, E. 2000. "Hume on Causality: Projectivist and Realist?" In *The New Hume Debate*, R. Read & K. A. Richman (eds), 113–21. London: Routledge.
- Davidson, D. 1967. "Causal Relations", *Journal of Philosophy* 64: 691–703. Reprinted in Sosa & Tooley (1993), pp. 75–87.
- Dowe, P. 2000. *Physical Causation*. Cambridge: Cambridge University Press.
- Dretske, F. I. 1977. "Laws of Nature", *Philosophy of Science* 44: 248–68.
- Dretske, F. I. 1985. "Review of *What is a Law of Nature?* by D. M. Armstrong", *The British Journal for the Philosophy of Science* 36: 79–81.
- Ducasse, C. J. 1968. *Truth, Knowledge and Causation*. London: RKP.
- Ducasse, C. J. 1969. *Causation and Types of Necessity*. New York: Dover.
- Earman, J. 1984. "Laws of Nature". In D. M. Armstrong, R. J. Bogdan (ed.), 191–223. Dordrecht: D. Reidel.
- Earman, J. 1993. "In Defence of Laws: Reflections on Bas van Fraassen's *Laws and Symmetry*", *Philosophy and Phenomenological Research* 53: 413–19.
- Earman, J. & J. Roberts 1999. "*Ceteris Paribus*, There is no Problem of Provisos", *Synthese* 118: 439–78.
- Ells, E. 1991. *Probabilistic Causation*. Cambridge: Cambridge University Press.
- Ehring, D. 1997. *Causation and Persistence*. Oxford: Oxford University Press.
- Elder, C. L. 1994. "Laws, Natures and Contingent Necessities", *Philosophy and Phenomenological Research* 54: 649–67.
- Ellis, B. D. 1999. "Causal Powers and Laws of Nature". In *Causation and Laws of Nature*, H. Sankey (ed.), 19–34. Dordrecht: Kluwer.
- Ellis, B. D. 2000. "Causal Laws and Singular Causation", *Philosophy and Phenomenological Research* 61: 329–51.
- Ellis, B. D. 2001. *Scientific Essentialism*. Cambridge: Cambridge University Press.
- Ellis, B. D. & C. E. Lierse 1994. "Dispositional Essentialism", *Australasian Journal of Philosophy* 72: 27–45.
- Fair, D. 1979. "Causation and the Flow of Energy", *Erkenntnis* 14: 219–50.
- Fales, E. 1990. *Causation and Universals*. London: Routledge.
- Falkenstein, L. 1998. "Hume's Answer to Kant", *Nous* 32: 331–60.
- Feigl, H. 1970. "The Orthodox' View of Theories: Remarks in Defence as Well as Critique". In *Minnesota Studies in the Philosophy of Science*, vol. 4, M. Radner & S. Winokur (eds), 3–16. Minneapolis: University of Minnesota Press.
- Fogelin, R. J. 1985. *Hume's Skepticism in the Treatise of Human Nature*. London: RKP.

- Friedman, M. 1974. "Explanation and Scientific Understanding", *Journal of Philosophy* 71: 5–19.
- Garfinkel, A. 1981. *Forms of Explanation*. New Haven, CT: Yale University Press.
- Garrett, D. 1993. "The Representation of Causation and Hume's Two Definitions of 'Cause'", *Noûs* 27: 167–90.
- Gasking, D. 1955. "Causation and Recipes", *Mind* 64: 479–87.
- Gillies, D. 2000. *Philosophical Theories of Probability*. London: Routledge.
- Glennan, S. S. 1996. "Mechanisms and the Nature of Causation", *Erkenntnis* 44: 49–71.
- Goodman, N. 1983. *Fact, Fiction and Forecast*, 4th edn. Cambridge, MA: Harvard University Press.
- Harré, R. 1970. "Powers", *The British Journal for the Philosophy of Science* 21: 81–101.
- Hausman, D. B. 1976. "Another Defence of the Deductive Model", *Southwestern Journal of Philosophy* 7: 111–17.
- Hausman, D. M. 1998. *Causal Asymmetries*. Cambridge: Cambridge University Press.
- Hausman, D. M. & J. Woodward 1999. "Independence, Invariance and the Causal Markov Condition", *The British Journal for the Philosophy of Science* 50: 521–83.
- Heathcote, A. & D. M. Armstrong 1991. "Causes and Laws", *Noûs* 25: 63–73.
- Hempel, C. G. 1965. *Aspects of Scientific Explanation*. New York: The Free Press.
- Hewstone, M. 1989. *Causal Attribution*. Oxford: Blackwell.
- Hitchcock, C. 1995. "Discussion: Salmon on Explanatory Relevance", *Philosophy of Science* 62: 304–20.
- Horwich, P. 1987. *Asymmetries in Time*. Cambridge, MA: MIT Press.
- Hume, D. 1739. *A Treatise of Human Nature*, L. A. Selby-Bigge & P. H. Nidditch (eds). Oxford: Clarendon Press, 1978.
- Hume, D. 1740. *An Abstract of A Treatise of Human Nature*, L. A. Selby-Bigge & P. H. Nidditch (eds). Oxford: Clarendon Press, 1978.
- Hume, David 1748. "An Enquiry Concerning Human Understanding". In *Enquiries Concerning Human Understanding and Concerning the Principles of Morals*, L. A. Selby-Bigge & P. H. Nidditch (eds). Oxford: Clarendon Press, 1974.
- Humphreys, P. 1981. "Aleatory Explanations", *Synthese* 48: 225–32.
- Jeffrey, R. 1969. "Statistical Explanation vs Statistical Inference". In *Essays in Honour of Carl G. Hempel*, N. Rescher (ed.), 104–13. Dordrecht: D. Reidel.
- Kant, I. 1787. *Critique of Pure Reason*, Norman Kemp Smith (trans.). New York: St Martin's Press, 1965.
- Kemp Smith, N. 1941. *The Philosophy of David Hume*. London: Macmillan.
- Keynes, J. M. 1921. *A Treatise on Probability*. London: Macmillan.
- Kim, J. 1971. "Causes and Events: Mackie on Causation", *Journal of Philosophy* 68: 429–41. Reprinted in Sosa & Tooley (1993), 205–7.
- Kim, J. 1973. "Causes and Counterfactuals", *Journal of Philosophy* 70: 570–2. Reprinted in Sosa & Tooley (1993), 60–74.
- Kim, J. 1993. *Supervenience and the Mind*. Cambridge: Cambridge University Press.
- Kim, J. 1999. "Hempel, Explanation and Metaphysics", *Philosophical Studies* 94: 1–20.
- King, G., R. O. Keohane, S. Verba 1994. *Designing Social Inquiry*. Princeton, NJ: Princeton University Press.
- Kitcher, P. 1976. "Explanation, Conjunction and Unification", *Journal of Philosophy* 73: 207–12.

- Kitcher, P. 1981. "Explanatory Unification", *Philosophy of Science* 48: 251–81.
- Kitcher, P. 1985. "Two Approaches to Explanation", *Journal of Philosophy* 82: 632–9.
- Kitcher, P. 1986. "Projecting the Order of Nature". In *Kant's Philosophy of Science*, R. E. Butts (ed.), 201–35. Dordrecht: D. Reidel.
- Kitcher, P. 1989. "Explanatory Unification and Causal Structure", *Minnesota Studies in the Philosophy of Science*, 13, 410–505. Minneapolis: University of Minnesota Press.
- Kitcher, P. & W. Salmon 1987. "Van Fraassen on Explanation", *Journal of Philosophy* 84: 315–30.
- Kline, D. A. 1985. "Humean Causation and the Necessity of Temporal Discontinuity", *Mind* 94: 550–6.
- Kneale, W. 1949. *Probability and Induction*. Oxford: Clarendon Press.
- Körner, S. 1953. "On Laws of Nature", *Mind* 62: 216–29.
- Kripke, S. 1972. *Naming and Necessity*. Oxford: Blackwell.
- Kripke, S. 1982. *Wittgenstein on Rules and Private Language*. Oxford: Blackwell.
- Lange, M. 1993. "Lawlikeness", *Noûs* 27: 1–21.
- Lange, M. 2000. *Natural Laws in Scientific Practice*. Oxford: Oxford University Press.
- Lewis, D. 1973. *Counterfactuals*. Cambridge, MA: Harvard University Press.
- Lewis, D. 1983. "New Work for a Theory of Universals", *Australasian Journal of Philosophy* 61: 343–77.
- Lewis, D. 1986a. *Philosophical Papers*, vol. II. Oxford: Oxford University Press.
- Lewis, D. 1986b. "Causal Explanation". See Lewis (1986a), 214–40.
- Lewis, D. 1986c. "Causation". See Lewis (1986a), 159–213.
- Lewis, D. 1986d. "Counterfactual Dependence and Time's Arrow". See Lewis (1986a), 32–66.
- Lewis, D. 1986e. "Postscripts to 'A Subjectivist's Guide to Objective Chance'". See Lewis (1986a), 114–32.
- Lewis, D. 1986f. "Introduction". See Lewis (1986a), ix–xvii.
- Lewis, D. 1999. *Papers in Metaphysics and Epistemology*. Cambridge: Cambridge University Press.
- Lewis, D. 2000. "Causation as Influence", *Journal of Philosophy* 97: 182–97.
- Lipkind, D. 1979. "Russell on the Notion of Cause", *Canadian Journal of Philosophy* 9: 701–20.
- Loewer, B. 1996. "Humean Supervenience", *Philosophical Topics* 24: 101–26.
- Lowe, E. J. 1998. *The Possibility of Metaphysics*. Oxford: Oxford University Press.
- Machamer, P., L. Darden, C. Craver 2000. "Thinking About Mechanisms", *Philosophy of Science* 67: 1–25.
- Mackie, J. L. 1966. "Counterfactuals and Causal Laws". In *Analytic Philosophy: First Series*, R. J. Butler (ed.), 65–80. Oxford: Basil Blackwell.
- Mackie, J. L. 1973. *Truth, Probability and Paradox*. Oxford: Oxford University Press.
- Mackie, J. L. 1974. *The Cement of the Universe: A Study of Causation*. Oxford: Clarendon Press.
- Mackie, J. L. 1977. "Dispositions, Grounds and Causes", *Synthese* 34: 361–70.
- Martin, C. B. 1994. "Dispositions and Conditionals", *The Philosophical Quarterly* 44: 1–8.
- McDermott, M. 1995. "Redundant Causation", *The British Journal for the Philosophy of Science* 40: 523–44.
- McGuigan, F. J. 1997. *Experimental Psychology*. Englewood Cliffs, NJ: Prentice Hall.

- Mellor, D. H. (ed.) 1978. *F. P. Ramsey, Foundations: Essays in Philosophy, Logic, Mathematics and Economics*. London: RKP.
- Mellor, D. H. 1991. *Matters of Metaphysics*. Cambridge: Cambridge University Press.
- Mellor, D. H. 1995. *The Facts of Causation*. London: Routledge.
- Mellor, D. H. 2000. "The Semantics and Ontology of Dispositions", *Mind* 109: 757–80.
- Menzies, P. 1989. "Probabilistic Causation and Causal Processes: A Critique of Lewis", *Philosophy of Science* 56: 642–63.
- Menzies, P. 1993. "Laws of Nature, Modality and Humean Supervenience". See Bacon *et al.* (1993), 195–225.
- Menzies, P. 1996. "Probabilistic Causation and the Pre-emption Problem", *Mind* 105: 85–117.
- Menzies, P. 1999. "Intrinsic Versus Extrinsic Conceptions of Causation". In *Causation and Laws of Nature*, H. Sankey (ed.), 313–29. Dordrecht: Kluwer.
- Menzies, P. & H. Price 1993. "Causation as a Secondary Quality", *The British Journal for the Philosophy of Science* 44: 187–203.
- Mill, J. S. 1911. *A System of Logic: Ratiocinative and Inductive*. London: Longmans, Green & Co.
- Molnar, G. 1969. "Kneale's Argument Revisited", *The Philosophical Review* 78: 79–89.
- Morrison, M. 1995. "Capacities, Tendencies and the Problem of Singular Causes", *Philosophy and Phenomenological Research* 55: 163–8.
- Mumford, S. 1998. *Dispositions*. Oxford: Clarendon Press.
- Oliver, A. 1996. "The Metaphysics of Properties", *Mind* 105: 1–80.
- Owens, D. 1992. *Causes and Coincidences*. Cambridge: Cambridge University Press.
- Papineau, D. 1985. "Probabilities and Causes", *Journal of Philosophy* 82: 57–74.
- Papineau, D. 1989. "Pure, Mixed, and Spurious Probabilities and Their Significance for a Reductionist Theory of Causation". *Minnesota Studies in the Philosophy of Science* 13, 307–48. Minneapolis: University of Minnesota Press.
- Popper, K. 1959. *The Logic of Scientific Discovery*. London: Hutchinson.
- Psillos, S. 1999. *Scientific Realism: How Science Tracks Truth*. London: Routledge.
- Psillos, S. 2002a. "Simply the Best: A Case for Abduction". In *Computational Logic: From Logic Programming into the Future*, F. Sadri & A. Kakas (eds), 602–25. Berlin and Heidelberg: Springer-Verlag.
- Psillos, S. 2002b. "Salt Does Dissolve in Water, but not Necessarily", *Analysis* 62: 255–7.
- Quine, W. V. 1974. *The Roots of Reference*. La Salle, IL: Open Court.
- Railton, P. 1978. "A Deductive-Nomological Model of Probabilistic Explanation", *Philosophy of Science* 45: 206–26.
- Railton, P. 1981. "Probability, Explanation and Information", *Synthese* 48: 233–56. Reprinted in *Explanation*, D.-H. Ruben (ed.), 160–81. Oxford: Oxford University Press, 1983.
- Ramsey, F. P. 1925. "Universals". See Mellor (1978), 17–39.
- Ramsey, F. P. 1928. "Universals of Law and of Fact". See Mellor (1978), 128–32.
- Ramsey, F. P. 1929. "General Propositions and Causality". See Mellor (1978), 133–51.
- Read, R. & K. A. Richman (eds) 2000. *The New Hume Debate*. London: Routledge.
- Redhead, M. 1990. "Explanation". In *Explanation and its Limits*, D. Knowles (ed.), 135–53. Cambridge: Cambridge University Press.

- Reichenbach, H. 1947. *Elements of Symbolic Logic*. New York: Macmillan.
- Reichenbach, H. 1956. *The Direction of Time*. Berkeley and Los Angeles: University of California Press.
- Richards, T. J. 1965. "Hume's Two Definitions of 'Cause'", *The Philosophical Quarterly* 15: 247–53. Reprinted in Chappell (1968), 148–61.
- Robinson, J. A. 1962. "Hume's Two Definitions of 'Cause'", *The Philosophical Quarterly* 12: 162–71. Reprinted in Chappell (1968), 129–47.
- Robinson, J. A. 1968. "Hume's Two Definitions of 'Cause' Reconsidered" See Chappell (1968), 162–8.
- Robison, W. L. 1977. "Hume's Causal Scepticism". In *David Hume: Bicentenary Papers*, G. P. Morice (ed.), 156–66. Austin, TX: University of Texas Press.
- Russell, B. 1918. "On the Notion of Cause". In *Mysticism and Logic*, B. Russell. London: Allen & Unwin.
- Russell, B. 1948. *Human Knowledge: Its Scope and Limits*. London: Routledge.
- Salmon, W. 1977. "Laws, Modalities and Counterfactuals", *Synthese* 35: 191–229.
- Salmon, W. 1984. *Scientific Explanation and the Causal Structure of the World*. Princeton, NJ: Princeton University Press.
- Salmon, W. 1985. "Conflicting Conceptions of Scientific Explanation", *Journal of Philosophy* 82: 651–4.
- Salmon, W. 1989. *Four Decades of Scientific Explanation*. Minneapolis: University of Minnesota Press.
- Salmon, W. 1997a. *Causality and Explanation*. Oxford: Oxford University Press.
- Salmon, W. 1997b. "Causality and Explanation: A Reply to Two Critics", *Philosophy of Science* 64: 461–77.
- Salmon, W., J. C. Jeffrey, J. G. Greeno 1971. *Statistical Explanation and Statistical Relevance*. Pittsburgh, PA: University of Pittsburgh Press.
- Schaffer, J. 2000. "Trumping Preemption", *Journal of Philosophy* 97: 165–81.
- Schlick, M. 1931. "Causality in Contemporary Physics". In *Philosophical Papers, Volume II*, H. L. Mudler & B. F. B. van de Velde-Schlick (eds), 176–209. Dordrecht: D. Reidel, 1979.
- Schlick, M. 1932. "Causation in Everyday Life and in Recent Science". In *Philosophical Papers, Volume II*, H. L. Mudler & B. F. B. van de Velde-Schlick (eds), 238–58. Dordrecht: D. Reidel, 1979.
- Scriven, M. 1958. "Definitions, Explanations and Theories". *Minnesota Studies in the Philosophy of Science*, 2, 99–195. Minneapolis: University of Minnesota Press.
- Scriven, M. 1962. "Explanations, Predictions and Laws". *Minnesota Studies in the Philosophy of Science*, 3. Minneapolis: University of Minnesota Press.
- Scriven, M. 1966. "Defects of the Necessary Condition Analysis of Causation". In *Philosophical Analysis and History*, W. Dray (ed.). New York: Harper and Row. Reprinted in Sosa & Tooley (1993), 56–9.
- Scriven, M. 1975. "Causation as Explanation", *Noûs* 9: 3–16.
- Sklar, L. 1999. "The Content of Science, The Methodology of Science and Hempel's Models of Explanation and Confirmation", *Philosophical Studies* 94: 21–34.
- Skyrms, B. 1980. *Causal Necessity*. New Haven, CT: Yale University Press.
- Smart, J. J. C. 1993. "Laws of Nature as a Species of Regularities" See Bacon *et al.* (1993), 152–69.
- Sosa, E. & M. Tooley (eds) 1993. *Causation*. Oxford: Oxford University Press.
- Stove, D. 1965. "Hume, Probability and Induction", *Philosophical Review* 74. Reprinted in Chappell (1968), 187–212.
- Strawson, G. 1989. *The Secret Connexion*. Oxford: Clarendon Press.

- Stroud, B. 1977. *Hume*. London: Routledge.
- Stroud, B. 1978. "Hume and the Idea of Causal Necessity", *Philosophical Studies* 33: 39–59.
- Suppes, P. 1984. *Probabilistic Metaphysics*. Oxford: Blackwell.
- Swoyer, C. 1982. "The Nature of Natural Laws", *Australasian Journal of Philosophy* 60: 202–23.
- Tooley, M. 1977. "The Nature of Laws", *Canadian Journal of Philosophy* 7: 667–98.
- Tooley, M. 1984. "Laws and Causal Relations", *Midwest Studies in Philosophy* 9: 93–112.
- Tooley, M. 1987. *Causation: A Realist Approach*. Oxford: Clarendon Press.
- Tooley, M. 1990. "Causation: Reductionism vs Realism", *Philosophy and Phenomenological Research* (supplement) 50: 215–36.
- Tooley, M. 1997. *Time, Tense and Causation*. Oxford: Clarendon Press.
- van Fraassen, B. C. 1980. *The Scientific Image*. Oxford: Clarendon Press.
- van Fraassen, B. C. 1989. *Laws and Symmetry*. Oxford: Oxford University Press.
- van Fraassen, B. C. 1993. "Armstrong, Cartwright, and Earman on *Laws and Symmetry*", *Philosophy and Phenomenological Research* 53: 431–44.
- von Wright, G. H. 1971. *Explanation and Understanding*. Ithaca, NY: Cornell University Press.
- von Wright, G. H. 1973. "On the Logic of the Causal Relations" See Sosa & Tooley (1993), 105–24.
- Wilson, F. 1999. *The Logic and Methodology of Science in Early Modern Thought*. Toronto: University of Toronto Press.
- Winkler, K. P. 1991. "The New Hume", *The Philosophical Review* 100: 541–79.
- Woodward, J. 1984. "A Theory of Singular Causal Explanation", *Erkenntnis* 21: 231–62. Reprinted in *Explanation*, D.-H. Ruben (ed.), 246–74. Oxford: Oxford University Press, 1993.
- Woodward, J. 1986. "Are Singular Causal Explanations Implicit Covering-Law Explanations?", *Canadian Journal of Philosophy* 2: 353–80.
- Woodward, J. 1989. "The Causal Mechanical Model of Explanation" *Minnesota Studies in the Philosophy of Science*, 13, 357–83. Minneapolis: University of Minnesota Press.
- Woodward, J. 1992. "Realism About Laws", *Erkenntnis* 36: 181–218.
- Woodward, J. 1997. "Explanation, Invariance and Intervention", *Philosophy of Science* 64 (proceedings): S26–41.
- Woodward, J. 2000. "Explanation and Invariance in the Special Sciences", *The British Journal for the Philosophy of Science* 51: 197–254.
- Wright, J. P. 1973. *The Sceptical Realism of David Hume*. Manchester: Manchester University Press.
- Ylikoski, P. 2001. *Understanding, Interests and Causal Explanation*. Academic Dissertation, Department of Moral and Social Philosophy, University of Helsinki.

Index

- accidentally true generalizations 8, 9, 137, 139, 142, 150, 151, 154, 171, 185, 271
 - and confirmation 143–4
 - and counterfactuals 145–6, 147, 184, 198, 201
 - and explanation 10–11, 222, 275–6
 - see also ADT view of laws; laws of nature; MRL view of laws; RVL
- Achinstein, P. 16
- Anscombe, G. E. M. 72–3, 78, 131
- AP-property 37ff.
- argument from causal verbs 72–4, 299
- Armstrong, D. M. x, 6, 13, 79, 138, 154, 161, 163, 164, 166–72, 189, 190, 207, 263, 266, 295, 296, 299, 304
 - on counterfactual theories of causation 100
 - criticism of MRL-view of laws 154–8, 303
 - on direct causal knowledge 167
 - on nomic singular causation 169–71
 - on uninstantiated laws 177
 - see also events; laws of nature; nomic necessitation; universals
- Armstrong–Dretske–Tooley (ADT) view of laws 13, 161–6, 173, 175, 176, 179, 198, 207, 208, 210, 303
 - and laws vs accidents 162, 171, 172, 211
 - as a non-reductive view 161–2, 163
 - vs RVL 161–3, 210–11
 - see also Armstrong; nomic necessitation; universals
- Ayer, A. J. 20, 142, 160, 161, 180, 303
- Barnes, E. 308
- Bealer, G. 304
- Beauchamp, T. 87, 297, 298, 299, 300, 301, 302
- Beebe, H. x, 305
- Bennett, J. 79
- Bird, A. x, 305
- Blackburn, S. 56
- Braithwaite, R. 141–2, 161, 180
- Bromberger, S. 224, 229, 276, 306
- Broughton, J. 298
- Butterfield, J. 158
- capacities 114, 187, 189–90, 197
 - Cartwright on 192–6
- Carnap, R. 3, 4, 20, 68, 143, 215, 216, 217–18, 246, 296, 303, 305; see also causation
- Carroll, J. W. 172, 175–7, 201, 303, 305; see also laws of nature
- Cartwright, N. x, 14, 73–4, 105, 114, 190–97, 278, 299, 300, 305, 307, 308
 - on the abstract/concrete distinction 73–4, 299
 - on *ceteris paribus* laws 191
 - on nomological machines 190–92
 - on observing causation 74
 - see also capacities
- causal field 61
- causal interaction 110, 115–16, 117, 121, 124, 126, 282
- causal laws 12, 77, 79, 105, 128, 138, 171, 174, 181, 194, 195, 215, 227, 282
- causal objectivism 23, 54
- causal priority 86–7, 91
- causal processes 13, 109, 110–12, 119–20, 301
 - and causal laws 282
 - and conserved quantities 121, 123,

- 124, 125
- vs events 110–11
- and mark-transmission 111–13, 117–19
- and material objects 111, 122
- and mechanistic theories of causation 111ff.
- and persistence 112–13, 122, 301
- vs pseudo-processes 111–12, 114–15, 118–19, 120, 122
- and regularities 121, 282
- vs spatiotemporal junk 119
- causal realism 22–3
 - as a non-reductive view 23, 130–31
 - and RVC 23
- causal statements
 - and events vs facts 79
 - extensional 76
 - general 68, 82
 - and laws 77, 223, 231
 - logical form of 75, 79
 - meaning of 81–2, 83–5
 - singular 75, 76, 77, 78, 86
 - truth-conditions of 77, 129, 130
- causation 3–8, 19ff., 57ff., 81ff., 107ff.
 - agency theory of 101–3, 129
 - Carnap on 68, 215–16, 217–18
 - vs causal explanation 306
 - conserved quantity theory of 13, 121ff.
 - and correlations 57–8, 62, 137, 216–17, 245, 254–5
 - counterfactual theories of 5, 13, 81–7, 92–101, 102–6, 120, 131–2
 - and counterfactuals 6, 13, 62, 71, 74, 82, 85–7, 92–6, 100, 102–6, 119–20, 121, 125–6
 - direction of 15, 98, 125, 230, 299, 301, 307
 - and evidence 6
 - and explanation 2, 6, 10–11, 14, 215–18, 220, 223, 224ff., 231–2, 233, 278, 281–2, 283, 284, 288
 - as an extrinsic relation 7, 84, 102, 128, 129, 130, 131, 132
 - generalist vs singularists accounts of 127–8
 - and human action 102–3
 - and Humean Supervenience 129–30
 - Humean vs non-Humean theories of 7–8, 15, 58, 77, 113, 128, 131–3
 - as an intrinsic relation 6–7, 84, 106, 128, 130, 131, 302
 - intrinsic vs extrinsic theories of 7, 128–9
 - intuitions about 6–7
 - as *imus* condition 4, 13, 87–92
 - and invariable sequence 59, 60, 61, 64, 66
 - and invariance 104, 106, 132
 - and invariant quantities 301
 - Kitcher on 278, 283, 288, 290–91
 - and laws 5, 61–2, 76–8, 89–90, 137–8, 169–71, 216, 233
 - Lewis on 93–4
 - and logical empiricism 4, 215–18
 - and manipulation 6, 101–2, 103, 233
 - and mark-transmission 118–20
 - mechanistic theories of 5, 107ff., 111ff., 132–3
 - as necessary and sufficient condition 4, 77, 88, 89
 - as necessity-in-the-circumstances 82
 - non-reductive accounts of 8, 118, 130–31
 - as an observable relation 5, 71–4, 131
 - persistence theories of 108–10, 111, 122, 301
 - platitudes of 6
 - and predictability 217–18
 - probabilistic 6, 14–15, 133, 246, 299, 308
 - process theory of 301
 - and recipes 6
 - reductive generalist accounts of 4, 8, 19, 130
 - reductive singularist accounts of 66, 130
 - reductive vs non-reductive accounts of 8, 129–31
 - as a secondary quality 102
 - and statistical explanation 246, 254–5
 - and supervenience 8, 129–31, 133
 - and temporal order 230
 - as a theoretical concept 5, 131
 - transference theories of 122
 - transitivity of 94
 - and unconditionality 62
 - see also* counterfactual dependence; Hume; Humeanism; IS model of explanation; nomic necessitation; pre-emption; RVC; Salmon; singular causation; SR model of explanation
- Clatterbaugh, K. 295
- Clendinnen, F. J. 300
- clinical trials 3, 105, 300
- Coffa, A. 251, 284
- Cohen, L. J. 148, 179–81, 204, 303; *see also* laws of nature
- common cause 254, 301
- Costa, M. J. 54
- counterfactual conditionals 5, 92, 114–15, 120, 145–6, 156, 206
 - active 184
 - backtracking 86, 98
 - and causal dependence 93–4

- and Dowe's conserved quantity theory 121, 125–7
- Goodman's theory of 147
- and interventions 104
- Lewis's theory of 92–3, 94–5, 130–31, 147–8
- Mackie's theory of 82–5, 131, 299
- and manipulation 101ff.
- and meaning of causal statements 81–2
- and possible worlds 82–3, 92–3
- and Salmon's theory of causation 114–15, 116, 117, 119–20
- see also* accidentally true generalizations; causation; laws of nature, MRL view of laws; RVL
- counterfactual dependence 92–6, 131, 238
 - and causal dependence 93–4
 - and causation 94, 97, 302
 - non-causal 100–101
 - see also* pre-emption
- Craig, E. 22, 54–5, 56, 297, 298
- Davidson, D. 2, 13, 57, 75–9, 85, 233, 299, 306
 - on singular causation and laws 76–8
 - see also* events
- deductive chauvinism 283–7
- deductive-nomological (DN) model of explanation 10, 14, 138, 215ff., 242, 249, 275, 281, 283
 - adequacy conditions of 220–1
 - and asymmetry of explanation 227, 229, 276
 - the basic thesis 222–4, 225, 226, 230–31, 232, 234, 237–8
 - and causal explanation 222–3, 228, 230, 231, 237, 238
 - counter-examples to 224ff., 231ff.
 - and false laws 307
 - and invariant relations 234
 - and laws 10, 222, 226ff.
 - and laws of co-existence 226–7, 228–9
 - and laws of succession 226–7, 229–30
 - and manipulation 228–9
 - and maximal specificity 249–50
 - and non-causal explanation 226, 227, 306
 - and prediction 235–6
 - see also* DNP model of explanation; IS model of explanation
- deductive-nomological-probabilistic (DNP) model of explanation 14, 256–61, 262, 286
 - and deductive-nomological explanation 256, 258, 259, 261
 - and ideal DNP text 260
 - and indeterminism 257
 - and inductive-statistical explanation 258–9
- deductive-statistical (DS) model of explanation 242, 284
- determinism 15, 207, 251, 285, 286
- dispositional essentialism 174, 305
- dispositions 114, 196–7, 305
 - functionalist theory of 197
 - Mackie on 196–7
 - rationalist theory of 196–7
 - realist theory of 196
 - see also* capacities
- Dowe, P. 13, 107, 116, 117, 119, 121–7, 301
 - on identity over time 123–4
 - on possession vs transmission of conserved quantities 124–5
 - see also* counterfactual conditionals; Salmon
- Dretske, F. 13, 156, 157, 161, 162, 163, 304; *see also* laws of nature
- Ducasse, C. J. 5, 12, 57–8, 66–73, 77, 82, 86, 127, 130, 138, 298
 - on causal verbs 72–3
 - on Hume's definition of causation 57–8, 68
 - on the observability of causation 71–2
 - on single difference 5, 66–71
 - see also* events; laws of nature; singular causation
- Earman, J. x, 152, 175, 176, 196, 197, 263, 303, 304, 305, 306
- Eells, E. 15
- effects of a common cause 62, 90–91, 236, 245–6
- Ehring, D. 122, 130, 300, 301, 302
- Elder, C. L. 173
- Ellis, B. D. 174, 197, 305
- epistemic fallacy 40
- the Euthyphro contrast 290
- events 2, 66–7, 79, 93, 110
 - Armstrong on 79
 - Davidson's theory of 75–6
 - Ducasse on 66–7
 - vs facts 79, 306
 - fragile 86, 299
 - Kim's theory of 79
 - and properties 79
 - as the relata of causation 66
 - see also* causal processes
- event-tokens vs event-types 67
- explanation 2, 10–12, 14, 215ff., 241ff., 263ff., 281ff.
 - actual vs potential 221
 - as arguments 219, 220, 232, 243, 245, 255, 259, 274–5

- asymmetries of 276–8, 308
- causal 222–4, 225, 226, 227–8, 229, 230, 231, 232, 233, 237, 238, 246, 288, 306
- and causal histories 236–7
- as causal stories 231
- of chance events 256–9, 261–2, 286–7
- contrastive 261–2
- covering law model of 222
- as deductive derivation 283–4
- epistemic conception of 219, 281, 282
- and explanatory information 237, 238, 239, 260
- Friedman on 265–72
- Kitcher on 272–9, 287
- and laws 10–11, 138, 177, 217–18, 222, 226–8, 232–3, 275–6, 282
- of laws 263ff.
- Lewis on 14, 236–9, 261–2
- of likely events 242ff.
- mechanistic view of 283
- the metaphysics of 281–3, 287–93
- modal conception of 281, 308
- as nomic expectability 218–9
- not arguments 12, 232, 255, 259
- ontic conception of 281–2
- pragmatics of 16, 307
- Salmon on 253–6, 282–3, 285–6
- of statistical regularities 241, 284, 285
- and understanding 218–19, 232, 259–60, 265
- and unification 11–12, 14, 224, 264, 265–6, 268, 270–71, 272–3, 283, 289
- of unlikely events 244–5, 256–9
- without laws 231–2, 307
- Woodward on 234–5
- see also* causation; DN model of explanation; DNP model of explanation; DS model of explanation; Humeanism; IS model of explanation
- explanatory dependence vs causal dependence 282–3, 287–93
- extrinsic relation 128
- Fair, D. 122
- Fales, E. 304
- Falkenstein, L. 297
- Feigl, H. 308
- Fogelin, R. J. 296, 297
- Friedman, M. 14, 224, 265–72, 279, 283, 308; *see also* explanation; unification
- Garfinkel, A. 16
- Garrett, D. 298
- Gasking, D. 295
- Gillies, D. 296, 308
- Glennan, S. 301
- Goldbach's conjecture 303
- Goodman, N. 142, 143, 144, 221, 303
- Harré, R. 305
- Hausman, D. B. 307
- Hausman, D. M. 13, 15, 79, 98, 101, 103–6, 132, 299, 306, 307, 308
- Heathcote, A. 79, 170
- Hempel, C. G. x, 10, 14, 78, 148, 218–24, 226–30, 232–3, 235, 241–6, 248–50, 264, 265, 275, 276, 283, 285, 305, 306; *see also* DN model of explanation; DS model of explanation; IS model of explanation
- Hewstone, M. 299
- Hitchcock, C. 122, 301
- Horwich, P. 15, 95, 96, 98, 100, 303, 307
- Hume, D. 4, 12, 16, 18ff., 57–8, 66, 68, 71, 77, 84, 107, 110, 111, 113, 116, 120, 160, 161, 215, 216, 295, 297, 298, 302
- as an advocate of RVC 19–21, 26, 47, 50, 52, 297
- the basic methodological maxim 24, 27, 40, 45, 54
- on causal beliefs 30, 41–3, 45
- on causal inference 27, 29–32, 35–6, 37, 41, 44, 46
- as a causal realist 22, 55
- on causation as an extrinsic relation 50–51
- on causation as philosophical and as natural relation 23–4, 26–7, 41–2, 44, 45, 48, 49–50, 52, 295
- on constant conjunction 21, 27–9, 31, 40, 43, 44, 47
- on contiguity and succession 21, 25
- a counterfactual definition of causation 81
- on custom or habit 37, 43, 44, 51
- on customary transition 43, 44
- on distinct ideas 30–31, 295
- double aspect view of causation 48, 52
- as an error-theorist 47, 52
- on felt determination 44, 45, 46, 50
- on imagination 40–43, 52
- on impressions 24, 44, 45, 46
- the manifestation thesis 39–40
- as a naturalist 297
- on the necessary connection (necessity) 4, 21, 26, 28–9, 32, 40–41, 44–8, 49, 50

- necessity₁ vs necessity₂ 32–40
- on powers 33, 39–40, 296
- on the Principle of Uniformity of Nature 30–32, 35–6, 38, 43
- as a projectivist 56
- on relations of ideas vs matters of fact 296
- as a sceptical realist 22
- and scepticism 31, 34
- on singular causation 49, 58–9
- theory of ideas 24–5, 54, 298
- on the traditional conception of Reason 34–6
- on the two definitions of ‘cause’ 25, 48–52, 297
- Humeanism
 - about causation 4–5, 8, 14, 19–20, 56, 61–2, 82, 107, 133, 137, 170–71, 288, 292–3
 - about explanation 14, 263, 265, 266, 268, 278, 284, 287, 288, 293
 - about laws 8–10, 13, 137–8, 139, 141, 145, 154, 161, 174, 175, 181, 197, 210, 292–3
- Humean Nomic Supervenience 175–7
- Humean Supervenience 129–30, 132, 133
- Humphreys, P. 308
- the identification problem 165
- indeterminism 15, 207, 252, 262, 285–6
- and deductivism 256ff.
- inductive-statistical (IS) model of explanation 14, 243ff., 253, 285
 - ambiguity of 248–50
 - and causal explanation 246
 - and causation 245–6
 - and DN explanation 243–4
 - and high probability 244, 245
 - and homogeneity 250, 251, 308
 - non-monotonic nature of 248
 - probability in 246–7
 - and reference classes 248–9
 - and the Requirement of Maximal Specificity 250–51, 308
 - and statistical generalisations 242, 244, 245, 246–7, 251
 - and unlikely events 244–5
 - see also* DNP model of explanation
- the inference problem 165, 166–7, 168, 304
- intervention 102, 104–5, 182–3, 185, 305
- intrinsic relation 128, 302
- Jeffrey, R. 245
- Kant, I. 292, 297
- Kemp Smith, N. 22, 26, 297
- Keynes, J. M. 38
- Kim, J. 79, 91–2, 100, 129, 219, 306, 307; *see also* events
- Kitcher, P. x, 14, 118, 119, 120, 251, 264, 268, 272–9, 281–2, 283–7, 288–92, 307, 308; *see also* causation; explanation; Salmon; unification
- Kline, D. 295
- Kneale, W. 145, 159, 160, 161, 164, 266, 303
- Körner, S. 303
- Kripke, S. 20, 59, 159, 161, 173
- Lange, M. x, 14, 141, 198–206, 305, 306; *see also* laws of nature
- lawlikeness 141, 143, 144, 149, 161, 204, 210
 - as an extrinsic feature 162, 163
 - as an intrinsic feature 151, 162, 188
- laws of nature 8–10, 61, 137ff., 159ff., 179ff., 216, 223, 282
- vs accidents 8–9, 13, 137, 139, 141ff., 145ff., 150–51, 154, 159, 172, 181–2, 184, 188, 198, 203–4, 205–6, 210, 211, 222, 235, 271, 275, 282
- agency view of 187–9
- Armstrong on 163, 164, 170, 171
- Carroll on 177, 201
- ceteris paribus* 191, 195, 305
- Cohen on 179–81
- and confirmation 141–2, 143, 199, 205–6
- contingent 9, 141, 152, 159, 172, 207, 208
- and counterfactuals 145–8, 156–7, 181–2, 184, 198, 201–3, 206, 207, 211, 302
- Dretske on 162
- Ducasse on 68
- and eliminative induction 179–80, 204
- and enumerative induction 179–80, 204
- fundamental 265, 268
- inference-ticket view of 181–2, 199, 306
- and initial conditions 160, 201
- intrinsic characterisation of 162–3, 188
- invariance of 13, 183–6, 188, 202–3
- Lange on 198–206
- Lewis on 149–50, 303
- logical empiricist view of 140, 216
- Mackie on 302
- Mellor on 207–10
- Menzies on 187–9
- as metaphysically necessary 10, 172–5, 207
- Mill on 61–2, 148–9

- modal force of 145–8
 and natural-kind predicates 140–41,
 144, 154–5, 271
 non-causal 171
 normative dimension of 199–200
 not knowable a priori 160
 and objective chances 207
 and prediction 142, 217–18
 Mill on 61–2, 148–9
 preservation of 202–3
 as principles of necessitation 159–60,
 266
 and projectability 144, 151
 Ramsey on 149–50, 151, 153–4, 303
 reductive vs non reductive views of
 161, 165, 167
 as relations among universals 9, 163–4
 stability of 13, 202–3
 Tooley on 163–4
 uninstantiated 143, 152, 164, 176,
 177, 210, 211
 Woodward on 182–7
see also ADT view of laws; causal laws;
 causation; DN model of explana-
 tion; explanation; Humeanism;
 MRL view of laws; necessity; nomic
 necessitation; properties; Ramsey-
 sentences; RVL
- Lewis, D. x, 5, 9, 13, 14, 21, 81, 92–
 101, 129, 147–8, 149–50, 155,
 236–9, 261–2, 299, 300, 302, 303,
 304
 on causal chains 94, 97
 on the direction of causal dependence
 97–8
 on the similarity among possible
 worlds 95–6
see also causation; counterfactual
 conditionals; explanation; laws of
 nature; pre-emption; properties;
 RVC
- Lierse, C. 174
 Lipkind, D. 295
 Loewer, B. 153, 176, 177, 303
 Lowe, E. J. 305
- Mach, E. 308
 Machamer, P. 301
 Mackie, J. L. x, 4, 13, 53, 74, 75, 79,
 81–92, 107–10, 112, 122, 131,
 196, 197, 297, 298, 299, 300, 302
 on causes as *inus* conditions 87–92
 critique of Hume 32–40, 84
 on the meaning of causal statements
 81–5
 on persistence 107–10, 301
see also causal priority; dispositions;
 laws of nature; RVC
- Martin, C. B. 305
- McDermott, M. 300
 McGuigan, F. J. 300
 mechanisms 13, 84, 91, 105, 259, 286
 causal 106, 107–8, 110ff., 132, 232,
 256, 260, 261, 282, 283, 289
 and persistence theories of causation
 107–10
see also causation; explanation
- Mellor, D. H. x, 6, 14, 79, 197, 207–
 10, 295, 299, 303, 305, 306, 308;
see also laws of nature; properties
- Menzies, P. 6, 13, 101, 102, 103, 129,
 187–9, 295, 300, 302, 303; *see also*
 laws of nature
- methods of agreement and difference
 12, 63–6
- Michotte, A. 299
- Mill, J. S. 4, 9, 20, 57, 58, 59–66, 71,
 87, 89, 137, 148–9, 298, 304
 on causes and conditions 60
 on positive and negative conditions
 59–60
 on unconditionality 62, 64–5
see also laws of nature; methods of
 agreement and difference, RVC
- Mill–Ramsey–Lewis (MRL) view of laws
 9, 13, 148–51, 171, 176, 209, 210,
 211, 223, 264, 283, 288
 and counterfactuals 151, 156–7
 and explanation of laws 177
 and laws vs accidents 150–51
 and natural properties 154–5, 172
 objective construal of 153–4, 291
 and uninstantiated laws 152, 163,
 176, 210
see also Armstrong; RVL; unification
- modularity 104–5, 300
- Molnar, G. 140
- Morrison, M. 195
- Mumford, S. 197, 305
- natural kinds 140, 144, 155, 172, 174,
 271
- necessity 4, 26, 48, 91, 107, 109, 113,
 177, 182, 205, 281
 a posteriori 161, 164
 directly observable 71–2
 and experience 30–32
 experimental 188
 Humean view of 9, 19, 21, 52, 133,
 139, 141, 159, 160, 161, 215, 216,
 303
 and laws of nature 9, 110, 139, 141,
 157–8, 159, 160, 161, 173–4, 182,
 203, 210
 logical 160, 161, 164, 173, 188
 metaphysical 10, 48, 159, 172–4, 305
 non-Humean view of 110, 159–61,
 172–5

- projective theory of 56
- and Reason 29–30
- see also* causation; Hume
- New Hume 12, 22, 52–6
 - and Hume as a causal realist 55
 - and Hume's theory of ideas 55
 - on supposing vs conceiving 54–5
- nomic necessitation 10, 163, 164–6, 166–72, 177, 208
 - Armstrong on 166–72, 209
 - as a causal relation 168, 169
 - contingent 159, 161, 164, 174, 304
 - and Humean regularities 164–5
 - and laws of nature 162–4
 - and singular causation 169ff.
 - as a theoretical construct 164
 - Tooley on 165–6, 304
 - and uninstantiated laws 163
 - as a universal 166
 - see also* the inference problem; universals
- Ockham's razor 40
- Oliver, A. 304
- Oppenheim, P. 218
- Orwell, G. 161
- overdetermination 85–7, 96, 299
- Owens, D. 300
- Papineau, D. x, 15, 308
- Peirce, C. S. 292
- plurality of causes 87
- Poincaré, H. 308
- Popper, K. 160, 161, 303
- Port-Royal Logic 304
- the possibility of divergence 289
- postulate of non-interference of nature 70
- power realism 54
- pre-emption 96–8, 300
 - and counterfactual dependence 99–100
 - and counterfactual theories of causation 97–8
 - early 99
 - late 99
 - Lewis on 96–8, 300
 - trumping 99–100
- Price, H. 13, 101, 102, 103, 129, 300
- principle of limited variety 38
- principle of uniformity of nature *see* Hume
- properties 119, 155, 171–2, 197, 207
 - and laws 145, 154–6, 172, 173–4, 207–10
 - Lewis on 155, 303
 - Mellor on 208–9
 - natural vs non-natural 155
 - as tropes 301
 - as universals 166–7, 172
 - see also* events; MRL view of laws; RVL
- Quine, W. V. 20, 304
- Railton, P. x, 14, 256–2, 286; *see also* DNP model of explanation
- Ramsey, F. P. x, 9, 143, 149–50, 151, 153–4, 181, 208, 292–3, 303, 306; *see also* laws of nature
- Ramsey-sentences 208
 - and laws 208–9
- Read, R. 298
- Redhead, M. 307
- Reichenbach, H. 111, 141, 144, 146, 230, 301
- regularity view of causation (RVC) 4, 12, 19–21, 53, 54, 56, 81, 87, 91, 92, 107, 130, 137–8, 170, 198
 - and causal inference 90
 - and causation as an extrinsic relation 128
 - and causes as *inus* conditions 88–9
 - and Lewis 92
 - Mackie's critique of 81, 84, 87, 107
 - Mackie's version of 87–9
 - Mill's defence and refinement of 12, 59–62
 - as an objective position 23, 54
 - and the regularity view of laws (RVL) 137–8
 - vs singular causation 68, 76–8, 127–8
 - and supervenience 130
 - see also* causal realism; Hume
- regularity view of laws (RVL) 8, 13, 14, 137–8, 145, 151, 159, 163, 179, 198, 210–11
 - and counterfactuals 145–8, 151
 - epistemic version of 141–5, 151, 206, 217
 - and laws vs accidents 8, 141–3, 145–7, 150–51, 154, 157, 161
 - naïve version of 139–40
 - and properties 145, 154–6
 - as a reductive view 161
 - and uninstantiated laws 151
 - the web-of-laws (MRL) version of 8–9, 148–54, 155, 156, 157, 260
 - see also* ADT view of laws; MRL view of laws; RVC
- Richards, T. J. 297
- Richman, K. A. 298
- Roberts, J. 305
- Robinson, J. A. 295, 297
- Robison, W. L. 50
- Rosenberg, A. 87, 297, 298, 299, 300, 301, 302
- Russell, B. 3, 20, 113, 295, 301

- Salmon, W. x, 5, 12, 13, 14, 107, 110–21, 123, 252–6, 281–3, 284–7, 289–90, 301, 303, 307, 308
 and conserved quantity theory of causation 124
 and Dowe's theory of causation 124–5
 and Hume 110–11
 on interaction vs intersection 116–18
 vs Kitcher on explanation and causation 282–3, 284–7, 290–91
 on the mark-method 111–18
 on mark-transmission 113–15, 117–18
 and Russell's at-at theory of motion 113
 theory of causation 116–17
see also causal interaction; counterfactual conditionals; explanation; SR model of explanation
- Schaffer, J. 99, 100, 300
- Schlick, M. 181, 215, 216, 217, 305
- Schweder, R. x
- screening off 62, 254–5
- Scriven, M. 3, 78, 98, 231–2, 243, 299, 307
- Sellars, W. 200
- Simpson's Paradox 255, 308
- single-case probabilities (chances) 207, 247, 258
- singular causation 5, 7, 12, 49, 57, 58, 59, 66ff., 84, 121, 127, 169, 170, 189, 232, 238
 and causal explanation 232–3
 Ducasse on 5, 59, 66, 67–8, 127
 vs general 127–8
 as an intrinsic relation 68, 128, 132, 133
 and laws 68–9, 82, 128, 169–70
 mechanistic version of 121, 132
 and non-Humeanism 130–31
 observability of 71–4, 167, 304
 reductive accounts of 130
 and supervenience 130
see also Armstrong; Davidson; Hume; nomic necessitation; RVC; universals
- Sklar, L. 306, 307
- Skyrms, B. 305
- Smart, J. J. C. 303
- statistical-relevance (SR) model of explanation 14, 252–6
 and causal relevance 253
 and causation 254–5
- Stove, D. 296
- Strawson, G. 22, 23, 37, 54–5
- Stroud, B. x, 45, 46, 52, 72, 296, 297, 298, 299
- Suppes, P. 15, 308
- Swoyer, C. 173, 174
- Tooley, M. 13, 23, 130, 131, 161, 163–4, 165–6, 167, 168, 190, 303, 304
see also laws of nature; nomic necessitation
- the truth-maker principle 167
- unification 11, 177, 263ff., 283, 289, 308
 and atomic sentences 267–8, 269–71, 272
 and deductive systematisation 264, 268, 272–3, 274
 and explanatory schemata (argument patterns) 273–4, 276–7, 278
 and the ideal Hume corpus 288–91
 and natural-kind predicates 270–71
 and prediction 272
 and the problem of conjunction 267
 and the web-of-laws (MRL) view 11, 264, 268
see also explanation
- universals 9, 166–7, 168, 190, 207, 209, 266, 301, 304, 306
 Armstrong on 166–7
dictum de omni 304
 higher-order 166, 167
 naturalism about 170
 necessitating relations among 9, 13, 55, 161, 162–3, 168, 174, 207
 quasi- 171
 realism about 166, 170, 207
 and singular causation 169–70
 as truth-makers of laws 164, 167, 177
 uninstantiated 170, 304
- van Fraassen, B. C. 16, 157–8, 165, 169, 306, 307
- von Mises, R. 307
- von Wright, G. 101–2, 228, 229, 281
- Wilson, F. 304
- Winkler, K. 22, 55
- Wittgenstein, L. 161
- Woodward, J. 13, 78, 101, 113, 103–6, 132, 182–7, 189, 190, 196, 201, 234–5, 305, 307; *see also* explanation; laws of nature
- Wright, J. P. 22
- Ylikoski, P. 308